



# L'évaluation objective des dysfonctionnements de la voix et de la parole; 2e partie: les dysphonies

Bernard Teston

## ► To cite this version:

Bernard Teston. L'évaluation objective des dysfonctionnements de la voix et de la parole; 2e partie: les dysphonies. Travaux Interdisciplinaires sur la Parole et le Langage, 2001, 20, pp.169-232. hal-00173552

**HAL Id: hal-00173552**

**<https://hal.science/hal-00173552>**

Submitted on 20 Sep 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# L'ÉVALUATION OBJECTIVE DES DYSFONCTIONNEMENTS DE LA VOIX ET DE LA PAROLE

## 2<sup>ème</sup> PARTIE : LES DYSPHONIES

Bernard Teston

### Résumé

*L'évaluation de la qualité de la voix et la perception des causes de sa dégradation à travers différents indices acoustiques a toujours été la préoccupation clinique principale des phoniatres. Comme dans les autres disciplines médicales, ils ont été attentifs à toutes les techniques qui seraient susceptibles de leur donner des informations complémentaires, pour aider au diagnostic et évaluer les effets des traitements chirurgicaux ou médicamenteux ou les progrès d'une rééducation. Cependant, la voix étant par essence émise pour être entendue, l'évaluation subjective "à l'écoute" reste la référence face à des méthodes d'évaluation objective très variées, souvent mal fondées et à l'efficacité parfois contestée. Nous présentons dans cette étude différentes méthodes acoustiques et aérodynamiques que nous avons toutes appliquées et testées avec les systèmes EVA 1 et 2. Malgré quelques faiblesses, les évaluations objectives des dysphonies progressent en qualité et pertinence, poussées par la nécessité de disposer de bilans objectifs dans le but de répondre à des expertises, autant qu'aux attentes des patients qui imposent peu à peu l'utilisation de telles méthodes en association avec celles plus traditionnelles.*

### Abstract

*The evaluation of voice quality and the perception of the causes of its impairment through various acoustic cues has always been the main clinical concern of speech therapists. As in other medical disciplines, they have been particularly heedful of all techniques which might be capable of supplying more information as a diagnostic aid and to evaluate the effect of surgical or chemical treatment or the progress of re-education. Since the very purpose of the human voice is to be heard, subjective evaluation by listening tests remain the standard reference compared to a number of objective evaluation methods which have been proposed and which have often little justification and are of doubtful efficiency. In this study we present different acoustic and aerodynamic methods which have all been applied and tested with the EVA 1 and 2 systems. Despite some weaknesses, the objective evaluation of dysphonia is making both qualitative and quantitative progress, driven by the need for objective appraisals in response to expert evaluations and to the expectations of patients which tend to impose such techniques more and more in association with more traditional methods.*

Mots-clés : dysphonie, larynx, voix, évaluation vocale.

Keywords : dysphonia, larynx, voice, vocal evaluation.

## 4. L'évaluation des dysphonies

### 4.1. Les dysphonies

Nous avons mentionné précédemment que les dysarthries des coordinations pneumophoniques et de la musculature intrinsèque du larynx provoquaient des dégradations de la voix appelées dysphonies. Ces dysfonctionnements vocaux peuvent également être la conséquence de changements morphologiques de l'anatomie du larynx, essentiellement au niveau de la glotte, provoqués par des modifications de sa structure ; la génération d'un excès de tissus biologiques, ou à l'inverse, des manques anatomiques provoqués par des gestes chirurgicaux.

#### 4.1.1. Les dysphonies d'origines neurologiques

Cette première famille regroupe les dysphonies provoquées par l'état neuromoteur du patient. La dysphonie peut être provoquée schématiquement par de multiples facteurs tels que l'hypotonie ou, a contrario, l'hypertonie de la musculature laryngée et respiratoire ou encore des tremblements, qui ont pour conséquence de moduler la hauteur, l'intensité et le timbre de la voix. Elle peut également être provoquée par un mauvais contrôle de la fermeture de la glotte, conséquence de spasmes ou de paralysies.

L'hypotonie a pour conséquence une faible intensité de la voix et un abaissement de la Fo (Baker *et al.*, 1998). L'hypertonie, qui se manifeste par la difficulté à initialiser un acte volontaire du larynx, se traduit par des hésitations au démarrage du voisement, des émissions vocales discontinues, une augmentation de la Fo, un timbre sourd par manque d'harmonique, et voilé par suite d'un mauvais accollement des cordes vocales (Stelzig *et al.*, 1999). Les tremblements, qui peuvent être de fréquence variable en fonction de leur origine, rendent la voix chevrotante (tremor), (Perez *et al.*, 1996) ; on peut leur associer des instabilités de Fo en voix tenue (Gentil, 1990).

Les dysphonies spasmodiques (ou dystonies laryngées) provoquent des changements brutaux de la hauteur de la voix qui peut s'interrompre, repartir, glisser et chevroter (tremor). Elle peut avoir un timbre désagréable et être, au pire, inintelligible.

Dans les paralysies laryngées, une corde vocale demeure en position plus ou moins ouverte à la suite d'un mauvais contrôle neuromoteur. La voix est monotone avec perte de mélodie et de nombreuses désonorisations. Elle est soufflée et rauque avec une fuite d'air importante, entraînant un essoufflement en fin de phrase et une voix projetée continue impossible.

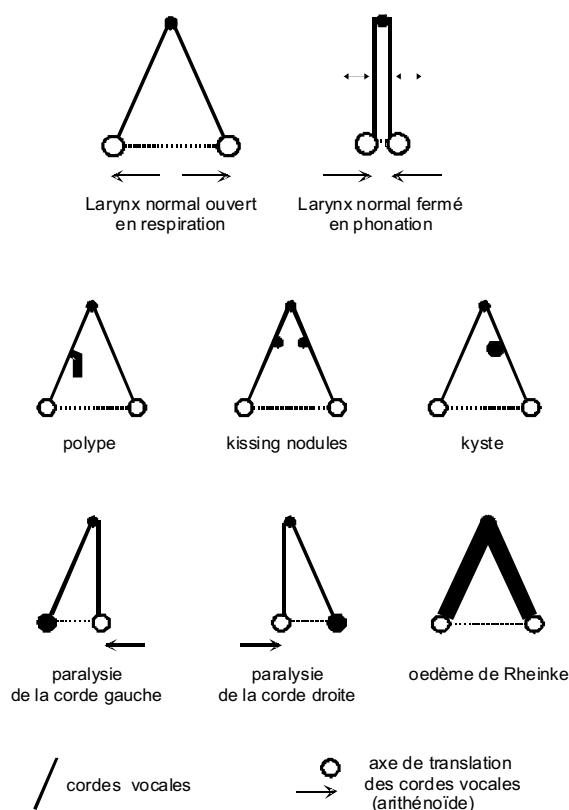
#### 4.1.2. Les dysphonies d'origines morphologiques

Cette seconde famille regroupe les changements anatomiques de la glotte provoqués par l'apparition de nodules, polypes et kystes qui sont des lésions bénignes des cordes vocales

provoquées généralement par un forçage vocal permanent ou brutal. La voix est plus grave, rauque, soufflée. Son timbre est voilé, sourd et éraillé.

Le changement de structure de la glotte peut être également provoqué par des laryngites qui sont des inflammations de l'ensemble des cordes vocales occasionnées par des infections, favorisées par l'effort vocal et qui peuvent s'installer de manière permanente (œdème de Reinke). La voix est plus grave, avec des difficultés dans les aigus, rauque et peu timbrée. Elle peut même disparaître totalement (extinction de voix).

Enfin, les plus importants changements anatomiques de la glotte sont provoqués par des traumatismes chirurgicaux à la suite de l'ablation d'un cancer cordal. La voix est très dégradée, quoique fonction de la technique chirurgicale, avec parfois des désonorisations. Elle est grave, de faible intensité, mais intelligible sauf dans le bruit. Le timbre est très rauque, granuleux, soufflé en rapport avec la fuite glottique.



**Figure 12**  
*Représentation schématique des atteintes morphologiques du larynx*

## 4.2. Les évaluations qualitatives

### 4.2.1. L'évaluation subjective à l'écoute

De nombreuses méthodes d'analyses perceptives ont été proposées pour l'évaluation de la qualité de la voix "à l'écoute" du patient par Hammarberg (1980), Hirano (1981), et Dejonckere (1993) entre autres. De toutes ces méthodes, c'est l'échelle GRBAS de Hirano (1981) qui semble la plus utilisée. C'est une échelle à quatre niveaux :

G définit le grade général de dysphonie ; 0 représente la voix normale, 1 une dysphonie légère, 2 une dysphonie moyenne et 3 une dysphonie sévère. R définit la raucité de la voix, B définit son souffle, A l'asthénie et S son forçage.

Ces quatre dernières caractéristiques de la voix sont notées sur la même échelle à quatre niveaux. En fait, si cette cotation est relativement aisée sur la raucité et le souffle, elle s'avère quasiment impossible sur l'asthénie et le forçage.

L'évaluation du grade général, même réduit à une échelle à quatre niveaux est entachée d'une variabilité importante inter et même intra auditeur. L'évaluation d'un échantillon de voix peut varier d'un niveau pour le même auditeur particulièrement éduqué, à différents moments, et dans les mêmes conditions surtout entre les niveaux extrêmes (0 et 1 ou 2 et 3).

La variabilité inter auditeur est encore plus importante par le fait qu'ils sont de culture et d'école cliniques différentes et que chaque phoniatre définit à l'usage, et parfois inconsciemment, ses propres critères subjectifs. Ce problème n'est pas propre à cette pratique mais il est général à toutes les évaluations auditives.

L'évaluation inter auditeur, à l'écoute absolue, c'est-à-dire sans référence de comparaison, peut atteindre plus de 50% d'erreur (Kreiman *et al.* 1993). Les scores s'améliorent énormément avec des jurys d'écoute (c'est-à-dire lorsque plusieurs auditeurs donnent simultanément une évaluation consensuelle) mais ces derniers sont très lourds à mettre en œuvre et impossibles en pratique clinique de routine. Les scores s'améliorent également si l'évaluation se fait en mode comparatif. Cette méthode est surtout valable pour la comparaison avant et après traitement. Pour être pratique et efficace, elle doit être réalisée en écoute comparative immédiate avec possibilité de passage instantané entre les échantillons de voix. Des programmes sur micro-ordinateurs existent déjà avec ces possibilités pour en faciliter la mise en œuvre.

Dans les cas d'évaluation primaire, la comparaison ne peut être faite qu'avec une référence. Le choix de cette dernière paraît bien difficile à établir. Là encore il semble être essentiellement individuel en dehors de tout exemple ou norme nécessairement marqués par la variabilité linguistique.

#### 4.2.2. La vidéolaryngoscopie

Il est inutile de présenter cette technique d'observation du larynx, aboutissement de la déjà longue évolution d'une lignée d'instruments d'exploration (Luckhaupt, 1997), dont les techniques vidéo ont permis depuis relativement peu de temps d'enregistrer les images. Elle permet d'avoir les informations qualitatives des cordes vocales, indispensables à tout bilan clinique et de les archiver. C'est l'outil privilégié du phoniatre qui lui permet de voir le larynx en respiration et en phonation.

Cependant cette méthode, malgré son importance dans l'observation morphologique, est mal adaptée à l'étude du fonctionnement laryngien. Elle donne très peu d'informations sur le fonctionnement du vibreur et cela malgré l'existence de la stroboscopie. En effet, cette dernière qui permet de décomposer les diverses phases d'un phénomène périodique, ne donnent de bonnes informations que sur un signal parfaitement stable. Elle est donc mal adaptée, par principe, à l'étude de larynx pathologiques dont les vibrations sont par essence atteintes d'irrégularités. Seule la vidéo rapide peut remplacer avec avantage la stroboscopie, mais la vitesse nécessaire à nos applications étant trop éloignée des 50 images par seconde de la télévision domestique, elle est pour l'instant très onéreuse et peu aisée en pratique clinique.

### 4.3. Les évaluations acoustiques objectives

#### 4.3.1. Les évaluations globales

On entend par évaluations acoustiques globales les évaluations qui donnent une indication générale sur les capacités de l'organe vocal, soit sur sa dynamique tonale et énergétique pour le phonétogramme (ou Voice Range Profile), soit sur son rendement pour le Temps Maximal de Phonation (TMP). Ces deux méthodes sont d'ailleurs les plus anciennes des méthodes d'évaluation objectives.

Le phonétogramme consiste à tracer l'étendue vocale dans un plan cartésien avec les fréquences en abscisse et les intensités en ordonnées, en portant sur un graphe les notes produites aux intensités les plus basses et les plus fortes possibles. En reliant les points mesurés, on obtient une sorte de patatoïde aux extrémités rétrécies, dont la face inférieure représente les valeurs d'intensité les plus faibles et la face supérieure les valeurs les plus fortes. La distance entre les deux extrémités représente la dynamique tonale (en notation musicale ou en Hz) et l'épaisseur du patatoïde la dynamique de l'intensité (en dB).

Le phonétogramme est présenté parfois pour la production vocale, comme l'équivalent de l'audiogramme pour l'audition. C'est une méthode très utilisée sous des formes variées par les phoniatres et les orthophonistes. Traditionnellement, ils utilisent un clavier pour produire les

notes que reproduisent les patients et mesurent l'intensité avec un sonomètre. Avec une bonne habitude, cette pratique est aisée pour les deux protagonistes tout en étant entachée par les biais de chaque praticien. La procédure du phonétogramme normalisé proposée par l'UEP est plus lourde bien qu'automatisée sur micro-ordinateur (Sulter *et al.*, 1994). Le patient a souvent des difficultés à reproduire précisément la note. Il peut se fatiguer la voix rapidement surtout dans des atteintes neurologiques. Cependant, cette procédure est plus précise et permet une comparaison plus fiable. Les informations données par le phonétogramme sont essentiellement qualitatives. Mais on peut le quantifier simplement sous la forme des coordonnées du centre de gravité du patatoïde associé à la dynamique vocale. Bien qu'un tel indice soit très réducteur, il donne une assez bonne indication de l'information contenue dans le phonétogramme.

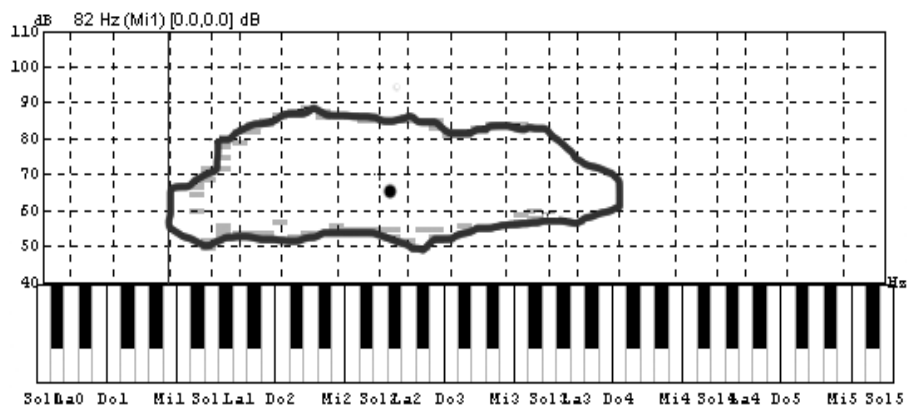
Il nous paraît à l'avenir nécessaire de simplifier la procédure du phonétogramme tout en conservant une norme, par exemple dans une évaluation en glissando. Bien que l'étendue vocale soit entachée par une variabilité individuelle importante, elle est bien corrélée avec l'état vocal du locuteur (Giovanni *et al.*, 1998). Le phonétogramme est à notre avis une investigation objective appelée à se développer à l'avenir pour l'analyse de l'état de certaines maladies neurologiques car il donne une représentation synthétique de la dynamique vocale référencée à sa valeur maximale possible et non à celle du discours.

La figure 13-A représente le phonétogramme d'un locuteur normal âgé de 58 ans. On constate une dynamique tonale du Mi 1 (82 Hz) au Do 4 (523 Hz) qui s'étend sur presque 3 octaves. La dynamique d'amplitude de l'émission vocale est de 40 dB. Son centre de gravité est sur le point de coordonnées 220 Hz 56 dB. On remarque un léger pincement à la hauteur du Do3, caractéristique du passage à la voix de fausset. Si cet exemple a une bonne dynamique tonale, il manque cependant un peu d'amplitude dans le médium.

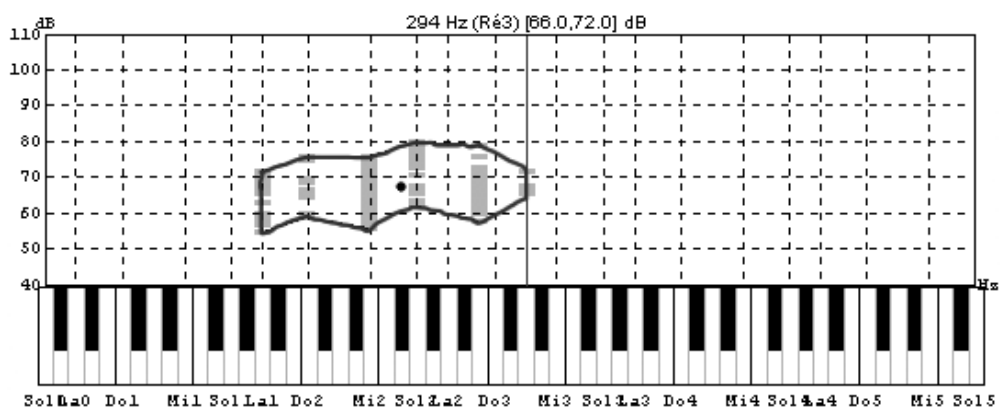
La figure 13-B représente le phonétogramme d'un patient parkinsonien. Sa dynamique tonale, bien inférieure à deux octaves, s'étend du La1 (110 Hz) au Ré3 (294 Hz) et sa dynamique d'amplitude atteint à peine 20 dB dans le meilleurs des cas. Cet écrasement de la dynamique est la conséquence de l'hypotonie caractéristique de son affection. Les coordonnées du centre de gravité de ce phonétogramme (196 Hz, 57 dB) sont très proches de celui de la figure précédente, malgré leurs grandes différences. Ceci montre que ces coordonnées doivent être obligatoirement associées aux valeurs des dynamiques tonales et d'amplitude pour caractériser un phonétogramme.

La figure 13-C représente un phonétogramme en voix parlée par un sujet normal, tracé sur la lecture du texte de la figure 10-A. Il ne doit absolument pas être comparé avec les deux

exemples précédents. Cette représentation, qui n'est pas normalisée, n'en apporte pas moins une bonne information sur la dynamique vocale naturelle d'un locuteur. Elle peut être considérée comme la synthèse des distributions des valeurs de  $F_0$  et d'amplitudes sur le signal de la phrase lue. Dans ce cas, on constate une dynamique tonale comprise entre Do2 (130 Hz) et La2 (220 Hz) avec une amplitude de 30 dB qui représente des valeurs normales dans la lecture sans accentuation prosodique particulière.

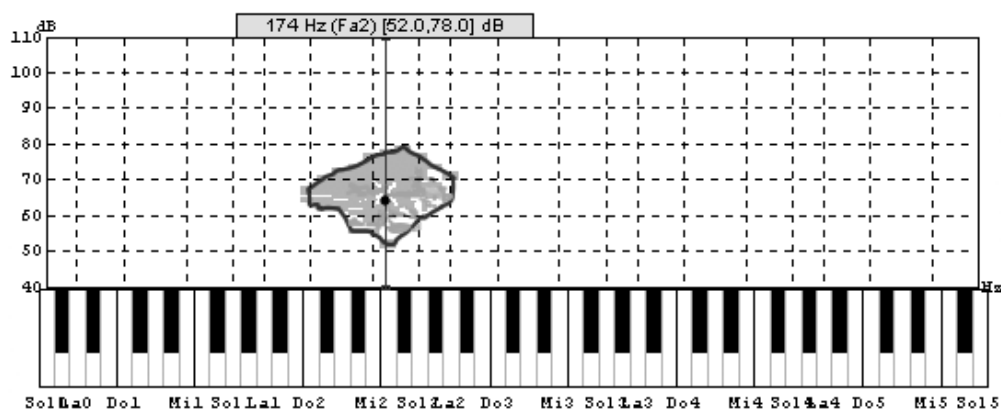


**Figure 13-A**  
*Phonétogramme d'un locuteur normal*



**Figure 13-B**  
*Phonétogramme d'un patient parkinsonien*





**Figure 13-C**  
*Phonétogramme d'un locuteur normal en voix parlée*

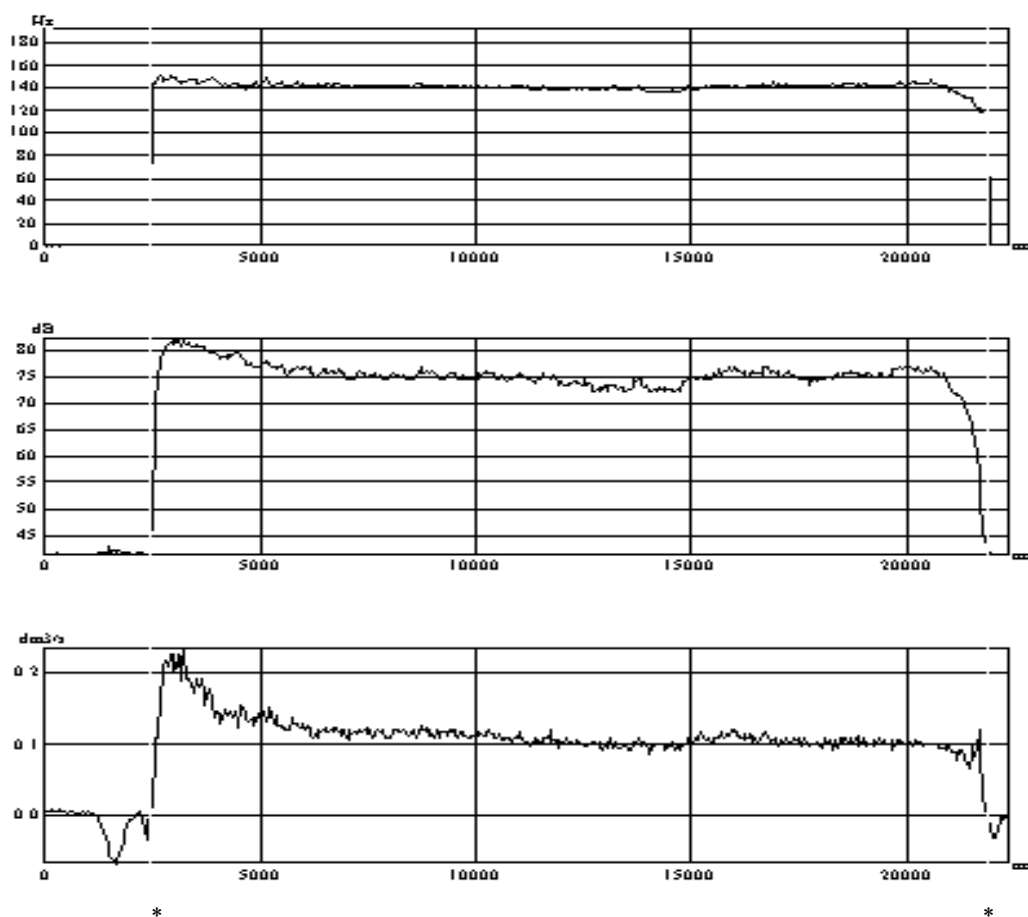
Le Temps Maximal de Phonation (TMP) consiste à mesurer le temps d'émission d'un "a" tenu, pour une hauteur et une intensité donnée. Il peut être mesuré à différentes hauteurs et intensité d'émission. Les valeurs usuelles sont généralement proposées. Traditionnellement, le phoniatre utilise pour ce faire un chronomètre associé ou non à un sonomètre. L'information du TMP s'apparente au rendement de la source vocale. En lui associant la mesure de la capacité vitale pulmonaire au moyen d'un spiromètre, on en déduit le quotient phonatoire (Hirano *et al.*, 1968).

Des procédures logicielles sur micro-ordinateurs permettent de mesurer précisément la fréquence et l'intensité vocale, certaines y associent la mesure du volume d'air, ce qui permet d'avoir une mesure précise du rendement vocal (voir mesures aérodynamiques).

La figure 14-A représente le TMP du sujet normal de la figure 13-A. On constate une durée d'émission de presque 20 secondes pour une  $F_0$  moyenne de 141 Hz et une intensité d'émission vocale de 75 dB, valeurs considérées comme normales pour un sujet masculin. Le volume total expiré de 2.22 dm<sup>3</sup> et le quotient phonatoire de 0.115 dm<sup>3</sup> par seconde sont également normaux. L'analyse des courbes de variation des paramètres montre la simultanéité de leur établissement, une bonne stabilité de la  $F_0$ , une légère instabilité à long terme sur l'intensité et un fort débit d'air au départ de la phonation qui devient très stable après quelques secondes. Le sujet a commencé sa production vocale avec une intensité qu'il a jugée trop forte pour durer longtemps. Il a alors changé rapidement de stratégie en diminuant l'intensité d'émission et donc le débit d'air. Ceci est un exemple de la variabilité du TMP en fonction de la stratégie du locuteur et montre le bien fondé de connaître les valeurs des paramètres acoustiques pour pondérer la valeur stricte du temps de phonation.

La figure 14-B représente le TMP d'un patient parkinsonien du même âge que le précédent. On constate un temps de phonation inférieur à 8 secondes, ce qui est court, pour une  $F_0$  moyenne de 167 Hz et une intensité d'émission vocale un peu faible de 70 dB. Le volume expiré est très faible ainsi que le quotient phonatoire, malgré un temps d'émission vocale très court. L'analyse des courbes de variation des paramètres montre une assez bonne stabilité à long terme de la  $F_0$  avec une déclivité régulière de 4 Hz par seconde environ. Il en est de même pour l'intensité avec une déclivité inférieure à 2 dB par seconde. Le phénomène de déclivité du débit est beaucoup plus marqué puisqu'il est divisé par 3 en moins de 8 secondes.

La figure 14-C représente le TMP d'une patiente de 45 ans atteinte par une paralysie laryngée. On constate une durée d'émission vocale inférieure à 10 secondes ainsi qu'un très difficile départ du voisement avec un retard important de plus d'une seconde par rapport au début du débit d'air oral. Les 3 premières secondes de l'émission vocale sont marquées par plusieurs désonorisations. Ensuite, malgré une bonne stabilité du débit oral, les vibrations du larynx restent instables et leurs valeurs déclinent de 250 à 160 Hz sur la durée d'émission jusqu'à l'extinction de la voix. Le quotient phonatoire de 0.235 dm<sup>3</sup>/s est très élevé. Il démontre une importante fuite glottique provoqué par un mauvais accolement des cordes vocales, qui est à mettre en relation avec la faible valeur de l'intensité de l'émission vocale (68 dB).

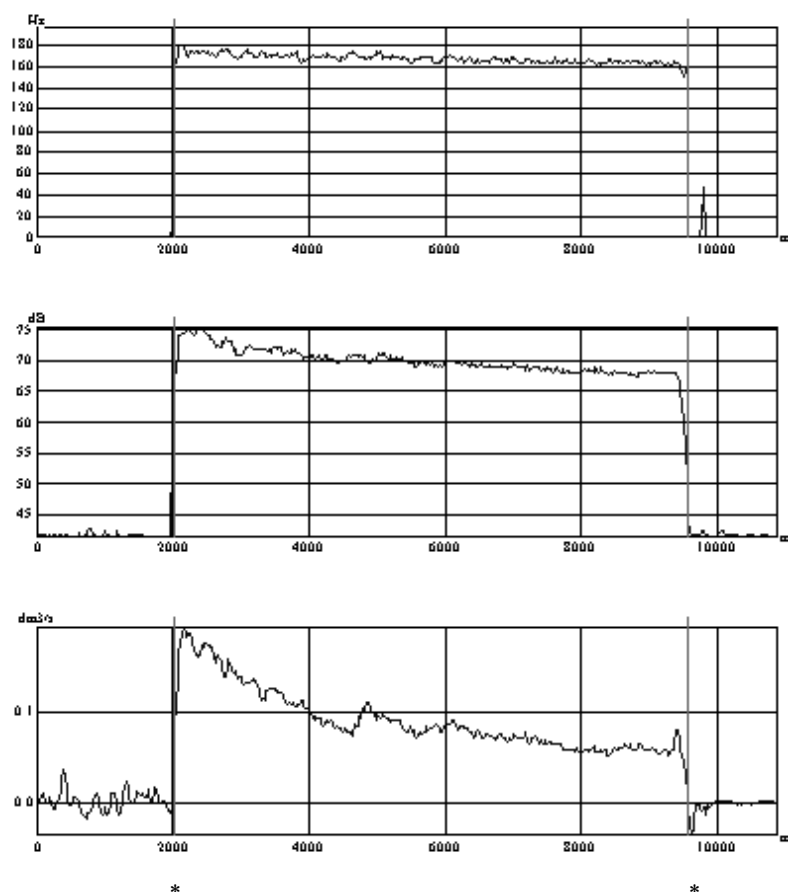


*Emission d'une voyelle "a" tenue le plus longtemps possible, avec de haut en bas :  
 les variations de la Fo ou mélodie en Hertz (Hz), celles de l'intensité en décibels (dB),  
 et celles de débit d'air oral en décimètre cube par seconde (dm<sup>3</sup>/s).  
 Les axes des abscisses représentent le temps en millisecondes (ms).*

Temps de phonation	Fo moyenne	Intensité moyenne	Volume expiré	Quotient phonatoire
19.33 s	141 Hz	75 dB	2.22 dm <sup>3</sup>	0.1149 dm <sup>3</sup> /s

*Tableau statistique sur la durée du temps de phonation mesuré  
 entre les deux curseurs (marqué par les étoiles \*)*

**Figure 14-A**  
*Temps maximal de phonation d'un sujet normal*

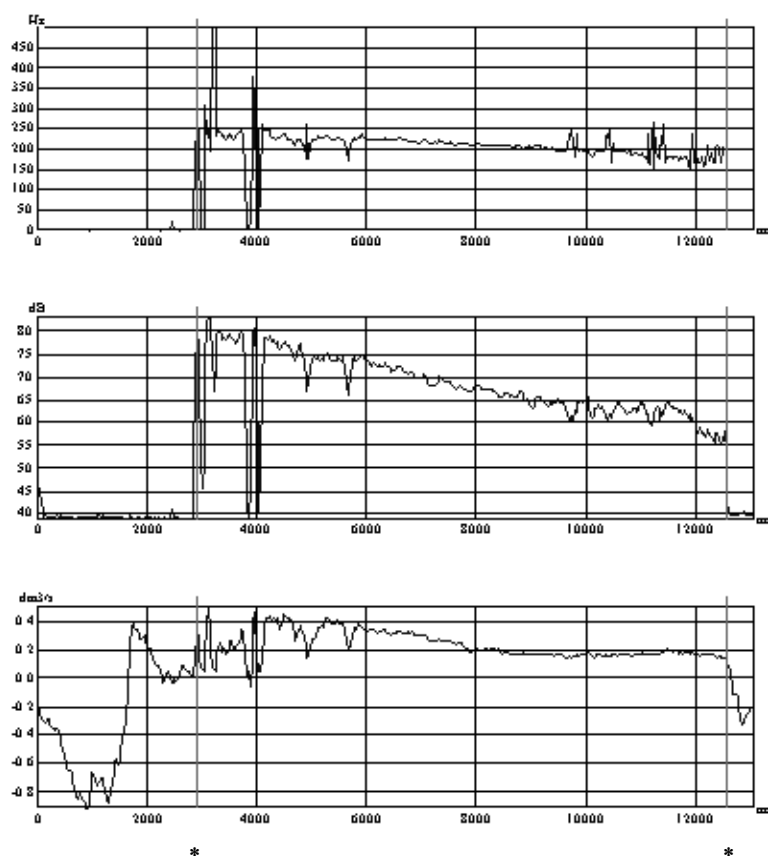


*Emission d'une voyelle "a" tenue le plus longtemps possible, avec de haut en bas :*  
*Courbe mélodique (Fo) (axe des ordonnées en Hertz [Hz]).*  
*Courbe d'intensité moyenne (RMS) du signal acoustique (axe des ordonnées en décibels [dB]).*  
*Courbe du débit d'air oral (axe des ordonnées en décimètre cube par seconde [dm³/s]).*  
*Les axes des abscisses représentent le temps en millisecondes (ms).*

Temps de phonation	Fo moyenne	Intensité moyenne	Volume expiré	Quotient phonatoire
7.37 s	167 Hz	70 dB	0.67 dm³	0.09 dm³/s

*Tableau statistique sur la durée du temps de phonation mesuré entre les deux curseurs  
(marqué par les étoiles \*)*

**Figure 14-B**  
*Temps maximal de phonation d'un patient parkinsonien*



*Emission d'une voyelle "a" tenue le plus longtemps possible, avec de haut en bas :*  
*Courbe mélodique (Fo) (axe des ordonnées en Hertz [Hz]).*  
*Courbe d'intensité moyenne (RMS) du signal acoustique (axe des ordonnées en décibels [dB]).*  
*Courbe du débit d'air oral (axe des ordonnées en décimètre cube par seconde [dm³/s]).*  
*Les axes des abscisses représentent le temps en millisecondes (ms).*

Temps de phonation	Fo moyenne	Intensité moyenne	Volume expiré	Quotient phonatoire
9.72 s	208 Hz	68 dB	2.29 dm³	0.235 dm³/s

*Tableau statistique sur la durée du temps de phonation mesuré entre les deux curseurs  
(marqué par les étoiles \*)*

**Figure 14-C**  
*Temps maximal de phonation d'une patiente souffrant d'une paralysie laryngée*

#### 4.3.2. Les évaluations acoustiques analytiques

Nous regroupons sous cette appellation toutes les évaluations autres que globales, c'est-à-dire qui s'attachent à faire ressortir des indices particuliers à partir des paramètres acoustiques plus ou moins porteurs d'informations sur les dysfonctionnements vocaux. On en dénombre plusieurs dizaines (Hirano, 1989), nous ne mentionnons que les principaux (Titze, 1995).

Si l'on se réfère à nouveau à l'échelle GRBAS, on peut considérer très schématiquement que la raucité est corrélée avec la stabilité du vibrateur laryngien, le souffle avec le rapport signal/bruit, l'asthénie avec l'intensité et le forçage avec une Fo plus haute.

##### 4.3.2.1. La stabilité du vibrateur laryngien

L'instabilité vibratoire de la glotte est une cause essentielle des dysphonies. Sa mesure est donc d'une grande importance dans leur évaluation. On l'effectue soit sur la fréquence soit sur l'amplitude du signal laryngien, à partir de la mélodie et de l'intensité de la voix. On peut définir plusieurs indices d'instabilité en fonction de la durée des fluctuations.

##### *Les fluctuations à court terme*

Les fluctuations à court terme, c'est-à-dire d'une durée de l'ordre d'un cycle glottique, caractérisent surtout les atteintes morphologiques des cordes vocales. Ces fluctuations sont le jitter pour la fréquence (Fo) et le shimmer pour l'amplitude (intensité). On peut également considérer la valeur statistique de l'écart-type comme un indice de la stabilité à court terme de la mélodie et de l'intensité.

La mesure du jitter pose de nombreux problèmes : il en existe plusieurs définitions mathématiques et sa valeur dépend de la technique de mesure de la fréquence fondamentale. Sur un même échantillon, elle n'est pas la même avec différents logiciels de mesure. Il existe plusieurs représentations du jitter :

Le « jitter absolu moyen » est la moyenne (sur une durée de l'ordre d'une seconde) de la différence de fréquence entre deux cycles vibratoires du larynx consécutifs. Ces variations de fréquence sont mesurées très précisément cycle à cycle.

Le « jitter factor » est le « jitter moyen » rapporté à la Fo moyenne du signal. C'est un bon indice pour évaluer la stabilité de la Fo à court terme. Il est considéré comme l'indice le plus significatif de la raucité de la voix.

Le « jitter ratio » est la moyenne de la différence de période entre deux cycles vibratoires consécutifs rapportée à la période moyenne du signal.

Enfin, le « RAP » est la moyenne de la différence de trois périodes consécutives, rapportée à la période moyenne du signal.

La mesure du shimmer est presque aussi problématique que celle du jitter. Là encore il existe plusieurs définitions du shimmer et de fortes différences entre les différents dispositifs.

Le « shimmer moyen » est la moyenne exprimée en dB (sur une durée de l'ordre d'une seconde) des différences d'amplitude entre deux cycles vibratoires du larynx consécutifs. Ces variations d'amplitude sont mesurées très précisément cycle à cycle.

Le « shimmer factor » est le « shimmer moyen » rapporté à l'amplitude moyenne du signal.

Enfin, l'APQ est la moyenne des variations d'amplitude sur 11 cycles vibratoires du larynx rapportée à l'amplitude moyenne du signal.

L'instabilité du signal glottique se manifeste comme un bruit qui lui est superposé. Elle peut donc également être évaluée au moyen du rapport entre l'énergie harmonique dans le spectre du signal et l'énergie du bruit. Il existe plusieurs méthodes pour mesurer la partie apériodique du signal vocal. Les méthodes du HNR (harmonic noise ratio ou rapport énergie des harmoniques/énergie du bruit qui s'exprime en dB) proposée par Yumoto *et al.* (1982) et la NNE (normalized noise energy ou énergie normalisée du bruit) proposée par Kasuya *et al.* (1986) en sont les principales mais leurs résultats ne sont pas uniquement fonction de la stabilité du signal glottique.

#### *Les fluctuations à moyen terme*

Les fluctuations à moyen terme, c'est-à-dire d'une durée maximale de l'ordre d'un phonème caractérisent plus généralement les atteintes neurologiques (tremor). Elles sont données par des calculs statistiques tels que le coefficient de variation (écart-type rapporté à la moyenne). Toutes ces fluctuations chiffrées peuvent être également représentées sous forme de distributions statistiques variées et c'est sous cette forme que leurs informations sont souvent plus directement accessibles. Dans beaucoup de cas, l'évolution de la fluctuation peut apporter plus de renseignements que sa valeur brute. C'est le cas en particulier des fluctuations contrôlées telles que le vibrato des chanteurs, dont la régularité montre qu'il ne s'agit pas d'un phénomène aléatoire propre aux dysfonctionnements (Ramig *et al.*, 1987).

On peut avancer qu'actuellement, les outils d'évaluation de la stabilité du vibrateur laryngien existent et fonctionnent de manière satisfaisante au sens de la mesure grâce à la puissance de la micro-informatique. Cependant, ils manquent d'efficacité clinique par absence surtout de normalisation, mais également d'exacte adaptation au problème. La moyenne du jitter ne rend pas bien compte de la variabilité instantanée du cycle glottique. Une méthode basée sur l'évaluation de la non linéarité du signal glottique au moyen du calcul du plus grand coefficient de Lyapounov (PGCL) est proposée dans ce sens avec des résultats prometteurs par Giovanni *et al.* (1999).

#### 4.3.2.2. Le souffle de la voix

C'est également un élément très important dans l'évaluation d'une dysphonie. Le souffle de la voix est considéré comme un bruit se superposant au signal vocal de la source laryngienne. L'analyse spectrale montre un spectre de raies bien défini pour un signal vocal de bonne qualité et un spectre continu massif pour un signal de bruit. Le rapport entre l'énergie du spectre harmonique et celle du spectre de bruit est le rapport signal sur bruit. Ce bruit peut être un bruit d'écoulement aérodynamique créé par une constriction du conduit vocal ou par un débit d'air trop important. C'est un bruit additif. Nous avons mentionné précédemment que le bruit peut être dû également à l'instabilité du signal glottique. Le HNR ou le NNE ne donnent donc, par principe, une bonne évaluation du bruit de souffle qu'avec une fréquence de vibration stable (comme pour la stroboscopie). Pour pallier ce défaut, Qi *et al.* (1995) ont proposé une méthode de mesure du HNR qui minimise les effets de l'instabilité vibratoire (jitter).

#### 4.3.3. Le corpus d'évaluation

Les évaluations mentionnées précédemment se font sur le phonème "a" tenu. Les avantages en sont multiples (De Krom, 1994 ; Giovanni *et al.*, 1996). C'est le phonème le plus spontané dans toutes les langues du monde (car le plus simple à articuler). Il minimise l'influence du conduit vocal sur la vibration du larynx. Il permet surtout d'évaluer la stabilité et le bruit du vibreur en régime permanent. Le problème essentiel de l'évaluation sur un "a" tenu est celui du choix de la fenêtre d'observation. En effet, la valeur des paramètres acoustiques mesurés peut varier dans de fortes proportions selon la durée de la fenêtre d'observation (nous préconisons une seconde) et son positionnement en début, en fin ou sur des parties plus ou moins stables de l'émission vocale. Nous préconisons également de la positionner sur la partie la plus stable (car la plus facile à juger), une seconde minimum après l'attaque.

Cette émission tenue de la voyelle "a" s'apparente à de la voix chantée. Or, une dysphonie paraît généralement plus légère en voix chantée qu'en voix parlée. Ceci est dû aux instabilités de vibration du larynx aux attaques vocaliques. Ces instabilités peuvent affecter toute la durée (assez courte) des noyaux vocaliques. Cependant ce sont là les conditions naturelles de fonctionnement en parole spontanée, mais dans lesquelles les évaluations développées précédemment ne peuvent plus s'appliquer.

#### 4.3.4. Les évaluations acoustiques des dysphonies en parole spontanée

Puisque les attaques vocaliques sont perturbées par les dysfonctionnements du larynx, on peut être tenté d'en tirer des informations sur leurs causes (Revis *et al.*, 1999). Ces informations peuvent être fournies par leur spectre acoustique. Il est possible d'analyser les attaques au moyen



de méthodes particulières telles que les ondelettes malgré la structure transitoire de leur signal pseudo stable. Les attaques doivent être auparavant segmentées, opération d'expertise longue et fastidieuse, pouvant difficilement être utilisée en pratique clinique de routine en l'état. Pour pallier ces inconvénients, il est possible également de tenter d'extraire les informations du signal de la source vocale à partir du spectre moyen à long terme d'un discours de l'ordre de la minute (Dejonckere, 1986). Des tentatives toutes récentes, utilisant des techniques proches de celles employées pour la reconnaissance du locuteur, cherchent à classer certaines atteintes pathologiques et leur importance. Enfin, les dysfonctionnements de la source vocale ont une influence sur la dynamique mélodique du discours. Une méthode initialement utilisée pour l'étude de la dysprosodie de patients neurologiques et décrite précédemment est en cours de test sur des dysphonies morphologiques.

#### 4.3.5. Les évaluations aérodynamiques

La voix et la parole sont la conséquence acoustique de phénomènes aérodynamiques qui se produisent dans le conduit vocal en fonction des mouvements des organes articulateurs. Les paramètres aérodynamiques sont au nombre de quatre : les pressions sous-glottique et intra-orale et les débits d'air nasal et oral. Seul ce dernier et la pression intra-orale sont impliqués dans l'évaluation pratique des dysphonies. Ils sont la manifestation des mécanismes physiologiques à l'origine de la production vocale. Ils sont mesurés par des méthodes simples moins fluctuantes que celles utilisées pour les paramètres acoustiques et reproductibles à condition de prendre soin aux conditions d'examen. Le débit d'air oral, associé aux paramètres acoustiques, permet de bien évaluer le rendement laryngien. Il permet également d'évaluer précisément la " fuite " glottique en phonation, en relation avec le bruit de souffle (Dejonckere, 1990), dans les cas où le complet accolement des cordes est empêché par des nodules, kystes ou autres polypes ou des paralysies laryngées. La connaissance du décours et des valeurs du débit d'air oral permet enfin de bien appréhender toutes les coordinations pneumo-phoniques (Jiang *et al.*, 1999). Il est possible de déduire aisément la valeur de la pression sous-glottique de celle de la pression intra-orale (PIO) au moyen de la méthode proposée par Smitheran *et al.* (1981). Associée au débit d'air oral, la pression sous-glottique permet de définir la résistance glottique comme étant le rapport de la PIO divisée par le débit d'air, hPa/(dm<sup>3</sup>/s) (Plant *et al.*, 1998). Parmi les indices proposés par Schutte (1980), l'efficacité glottique représentée par le rapport de l'intensité divisée par la PIO (dB/hPa) et le rendement glottique représenté par le même rapport

divisé par le débit d'air  $\text{dB}/[\text{hPa} \times (\text{dm}^3/\text{s})]$ , seul ce dernier est homogène au sens de l'énergie car les dB représentent l'énergie acoustique et la grandeur  $\text{hPa} \times (\text{dm}^3/\text{s})$  représente l'énergie aérodynamique. La figure 21 montre des exemples de mesures de ces différents indices qui se rapportent tous à l'évaluation de l'efficacité du larynx (ou laryngée, ou laryngienne).

La connaissance de la valeur de la pression sous-glottique à laquelle le forçage vocal est fortement corrélé (Giovanni *et al.*, 2000) a une grande importance en laryngologie, car il est responsable de la majorité des lésions morphologiques des cordes vocales. De ce point de vue, la résistance glottique nous paraît être l'indice le plus corrélé avec la force de contact des cordes vocales, donc du forçage vocal.

Les évaluations aérodynamiques sont d'une grande richesse mais restent malheureusement peu accessibles à cause de leur coût. Cela s'explique par le fait que les capteurs aérodynamiques sont plus complexes qu'un simple microphone.

#### 4.3.6. Méthodologie proposée pour l'évaluation des dysfonctionnements du vibreur laryngien

Comme pour les dysprosodies, nous avons développé des applications logicielles à partir des instruments de mesure des paramètres acoustiques et aérodynamiques des systèmes EVA 1 et 2, plus particulièrement dédiées à l'évaluation des dysphonies. Ces applications intègrent l'essentiel des indices pertinents des différents paramètres acoustiques et aérodynamiques décrits précédemment, dans un environnement et une ergonomie commune qui favorisent une utilisation clinique de routine. Nous avons fait en sorte de représenter, de la manière la plus synthétique possible, les variations qualitatives des paramètres  $F_0$ , intensité, débit et pression intra orale (PIO), en association avec des indices quantitatifs. Les décours de tous les paramètres sont donc visualisés de manière synchrone. Une fenêtre d'observation commune est alors choisie (généralement d'une durée d'une seconde), dans laquelle sont calculés tous les indices exposés dans le tableau 1. À ces valeurs numériques, sont associés des histogrammes qui présentent la distribution des paramètres, et des portraits de phase qui permettent de différencier les instabilités à court terme de celles à moyen terme.

Un diagramme polaire permet de synthétiser de manière qualitative l'importance relative de tous les indices mesurés. Cette représentation connue sous le nom de profil vocal (voice profile) a été proposée dans le MDVP (Multi-Dimensional Voice Program) par Kay Elemetric voici une dizaine d'années. Elle est qualifiée de polaire car elle utilise un système de coordonnées circulaire (polaire) où la place de chaque indice est donnée par un angle et son amplitude par la longueur d'un rayon. Le cercle de rayon maximum représente la limite supérieure de patho-

logie et le petit cercle la valeur des indices, frontière de normalité. Ces dernières ainsi que les limites supérieures de pathologie des indices présentés sur le profil vocal sont paramétrables. Elles ont, par défaut, les valeurs présentées dans le tableau 1. Elles ont été choisies soit dans l'ouvrage de Baken (1987) soit à partir de travaux menés au sein du laboratoire d'Audio-Phonologie clinique de Marseille avec lequel nous avons développé en collaboration les systèmes EVA.

Paramètre	Unités	Frontière normalité	Limite supérieure pathologique	Relation avec l'état pathologique
Débit d'air oral	dm <sup>3</sup> /s	0.2	0.5	++++
Pression sous glottique	hPa	7	18	++++
Efficacité glottique	dB/hPa	12.5	5	-----
Fuite glottique	cm <sup>3</sup> /s/dB	4	8	++++
Rendement glottique	dB/(hPa.dm <sup>3</sup> /s)	60	10	-----
Résistance glottique	hPa/dm <sup>3</sup> /s	40	100	++++
Jitter absolu	Hz	0.3	4	++++
Jitter Factor	o/o	0.99	5	++++
Jitter Ratio	o/oo	8	15	++++
C. V. Fo	o/o	1.5	4	++++
RAP	---	0.003	0.01	++++
Shimmer	dB	0.4	1.5	++++
Shimmer Factor	o/o	0.99	5	++++
APQ	---	0.1	0.3	++++
Signal Ratio	o/o	70	0	----
H/N Ratio	dB	7.4	-10	----
NNE	dB	-6	0	++++

*La notation +++ indique que la valeur du paramètre est en relation directe avec le grade de pathologie (plus la valeur du paramètre augmente, plus elle traduit un grade de pathologie élevé).*

*La notation --- indique que la valeur du paramètre est en relation inverse avec le grade de pathologie (plus la valeur du paramètre diminue, plus elle traduit un grade de pathologie élevé)*

### **Tableau 1**

*Les différents indices d'évaluation des dysphonies du système EVA 2, calculés à partir des paramètres acoustiques et aérodynamiques*

La figure 15-A représente de haut en bas : les variations des paramètres Fo, intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée de 6 secondes par une patiente souffrant d'un kyste cordal d'un volume important (figure 15-D), qui provoque une forte dysphonie de grade 2 selon l'échelle GRBAS. La voix est rauque et enrouée.

L'étude du décours de ces trois paramètres montre une bonne stabilité générale et donc l'absence de problèmes de contrôle pneumo-phonique. On constate une instabilité particulière de la fréquence fondamentale à court et moyen terme mais avec une certaine régularité. Les petits accidents impulsifs sur la courbe du débit d'air oral sont la conséquence d'un encombrement liquide de la glotte. Ils sont la manifestation physique de l'explosion de bulles d'air dans du mucus recouvrant les cordes vocales. Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités. Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1600 et 2600 millisecondes.

La moyenne de la Fo de 215 Hz est normale pour une voix de femme, par contre son écart-type de 8.4 Hz, son jitter (factor) de 4.6% et son coefficient de variation de 4% confirment une importante instabilité de la Fo entrevue à la lecture des courbes. Ces deux dernières valeurs sont d'ailleurs aux limites supérieures de la pathologie (Tableau 1).

La moyenne de l'intensité de 90 dB montre une voix forte, son shimmer de 0.6% se situe un peu au dessus de la frontière de normalité.

La moyenne du débit d'air est de 0.069 dm<sup>3</sup>/s, valeur normale pour une voix de femme. Sa variabilité est également normale pour un débit. Un indice de rendement laryngien est calculé comme la quantité d'air utilisée par dB d'énergie de voix émise. Il est exprimé en cm<sup>3</sup>/dB. Moins le larynx utilise d'air pour une intensité donnée, plus faible est cet indice et meilleur est son rendement. Il est également connu sous l'appellation de fuite glottique qui a pour avantage une relation directe entre sa valeur et le niveau de pathologie. Dans ce cas, la valeur de la fuite glottique est de 0.774 cm<sup>3</sup>/dB, ce qui est faible et démontre un bon rendement vocal. Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation. On constate une faible émergence harmonique avec une forte composante de « bruit ». Ce dernier est un bruit d'instabilité à cause du fort jitter et non un bruit de souffle, à cause du faible débit d'air oral moyen. Un tel spectre dénote une voix manquant de clarté, rauque, une voix enrouée.

La figure 15-B contient de haut en bas : les distributions des variations des paramètres Fo, intensité et débit sous la forme d'histogrammes et de portraits de phase. Cette dernière représentation graphique montre sur la diagonale l'instabilité à moyen terme du paramètre, sur la durée de la fenêtre d'observation. Son instabilité à court terme est représentée par l'impor-

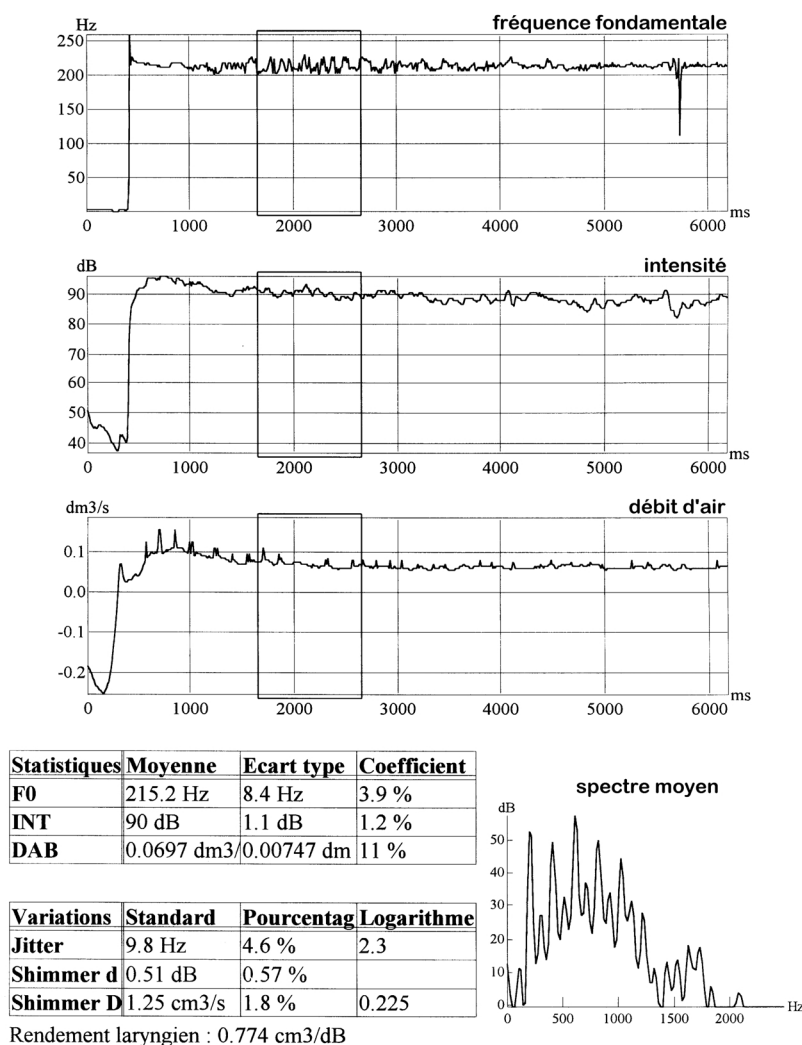
tance de la nuée des deux côtés de la diagonale. Si le paramètre est stable dans le temps, sa valeur apparaît sous la forme d'un point au centre du portrait de phase.

L'histogramme de la  $F_0$  montre une bimodalité très forte de la distribution des vibrations de la glotte. Ce phénomène est le signe évident d'un mauvais couplage des cordes qui vibrent différemment. Dans ce cas, la différence est de l'ordre de 15 Hz. La corde gauche, chargée par la masse du kyste (figure 15-D) vibre avec une fréquence inférieure à celle de la corde droite et ceci est la cause de l'importante raucité de la voix. Ce phénomène est confirmé par l'analyse du portrait de phase de la  $F_0$  ou l'instabilité à court terme apparaît sous la forme de deux nuées de part et d'autre de la diagonale.

La distribution de l'intensité est étroite et, comme sa valeur moyenne, elle est peu affectée par la présence du kyste. Son portrait de phase montre une faible instabilité à moyen terme sur la diagonale et une très faible instabilité à court terme. Cet exemple montre que, contrairement à un point de vue assez général, tous les paramètres acoustiques peuvent ne pas être affectés avec la même importance par une forte perturbation.

La distribution du débit d'air oral est normale si l'on tient compte des accidents impulsifs provoqués par les bulles de mucus.

La figure 15-C représente le profil vocal qui synthétise schématiquement tous les indices mesurés à partir des paramètres  $F_0$ , intensité et débit. On remarque que l'essentiel du dysfonctionnement glottique affecte, au-delà du cercle marquant la frontière de normalité, les quatre indices de stabilité vibratoire.



De haut en bas : Les variations de la mélodie ou Fo en Hertz (Hz), celles de l'intensité de la voix en décibel (dB), et celles du débit d'air oral en centimètre cube par seconde (cm<sup>3</sup>/s).

L'axe des abscisses est en millisecondes.

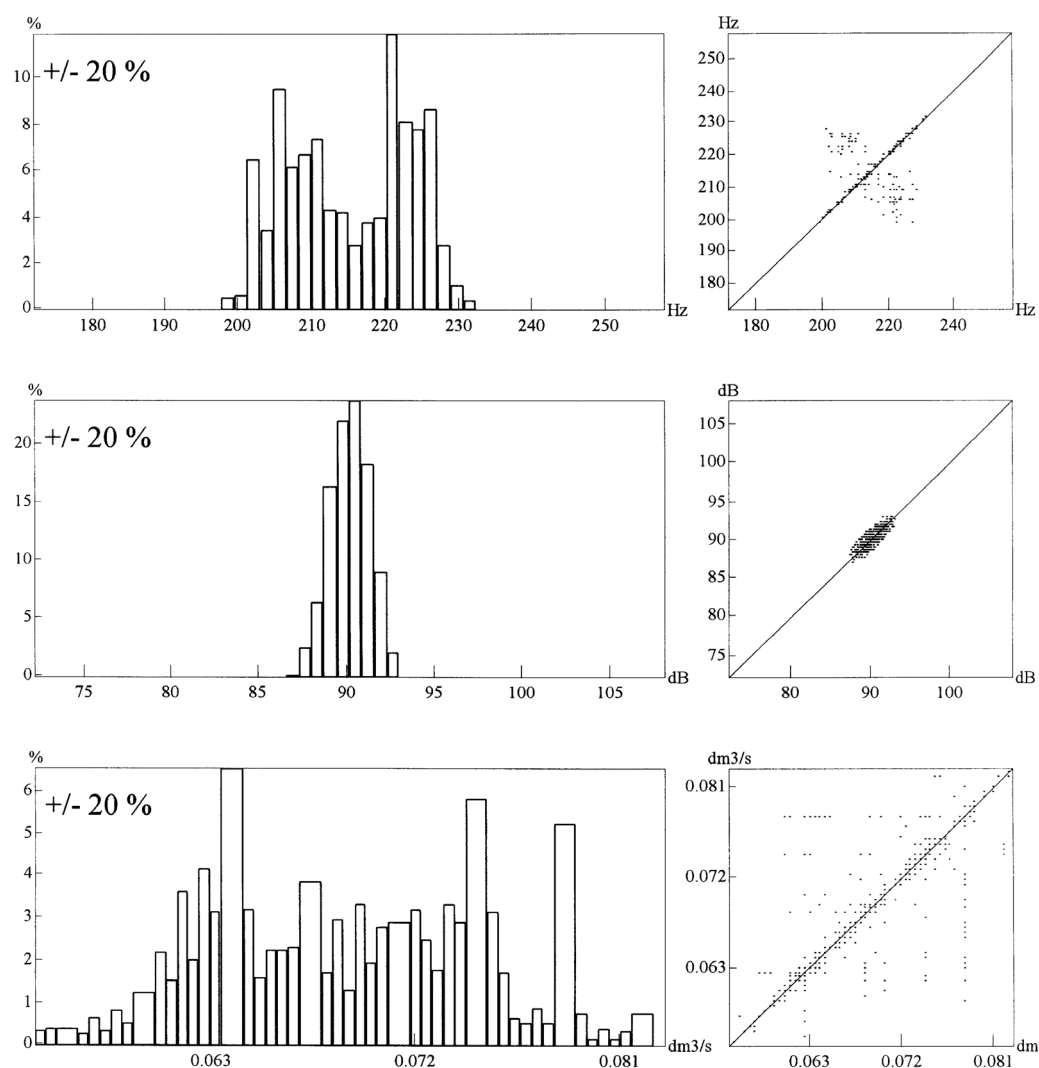
Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités.

Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1600 et 2600 millisecondes.

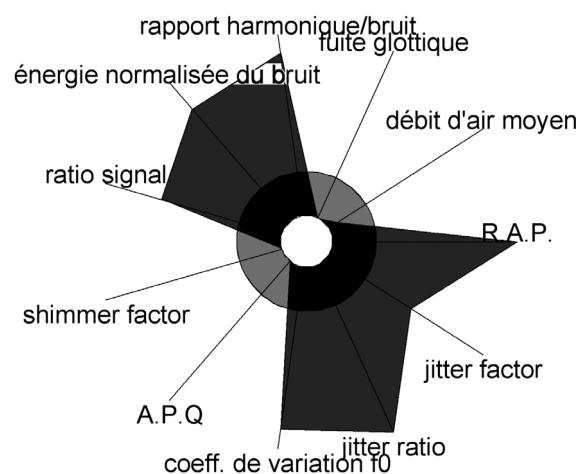
Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation.

### Figure 15-A

*Evaluation vocale sur la production d'un « a » tenu, par une patiente souffrant d'un kyste cordal*



**Figure 15-B**  
*Distribution et portraits de phase des variations des paramètres  $F_0$  intensité et débit d'air oral du cas précédent*



**Figure 15-C**

*Profil vocal de l'état de dysphonie à travers l'amplitude des différents paramètres du cas précédent*



*Cette image de la glotte, ayant produit le « a » tenu étudié précédemment, montre un kyste sur la corde gauche (d'un diamètre de l'ordre de 5 mm) qui occupe approximativement 1/3 de sa longueur. Il semble être situé sur la face supérieure de la corde et ne pas trop gêner la fermeture de la glotte.*

*Mécaniquement, il se comporte comme un balourd qui perturbe la vibration de la corde.*

**Figure 15-D**

*Photolaryngoscopie de la glotte du cas précédent*



La figure 16-A représente de haut en bas : les variations des paramètres Fo, intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée de 4 secondes par un locuteur âgé de 58 ans dont la voix est jugée normale (le même que pour le TMP de la figure 14-A).

L'étude du décours de ces trois paramètres montre une bonne stabilité générale et l'absence de problèmes de contrôle pneumo-phonique. On constate une instabilité transitoire au démarrage de la Fo très vite contrôlée. La courbe du débit d'air oral montre une légère déclivité régulière de l'ordre de 50 cc3/s signe d'un faible volume courant respiratoire dans ce cas. Les mesures statistiques sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1250 et 2250 millisecondes.

La moyenne de la Fo de 106 Hz est normale pour une voix d'homme, son écart-type de 0.9 Hz, son jitter (factor) de 0.32 % et son coefficient de variation de 0.8 % confirment la bonne stabilité de la Fo entrevue à la lecture des courbes.

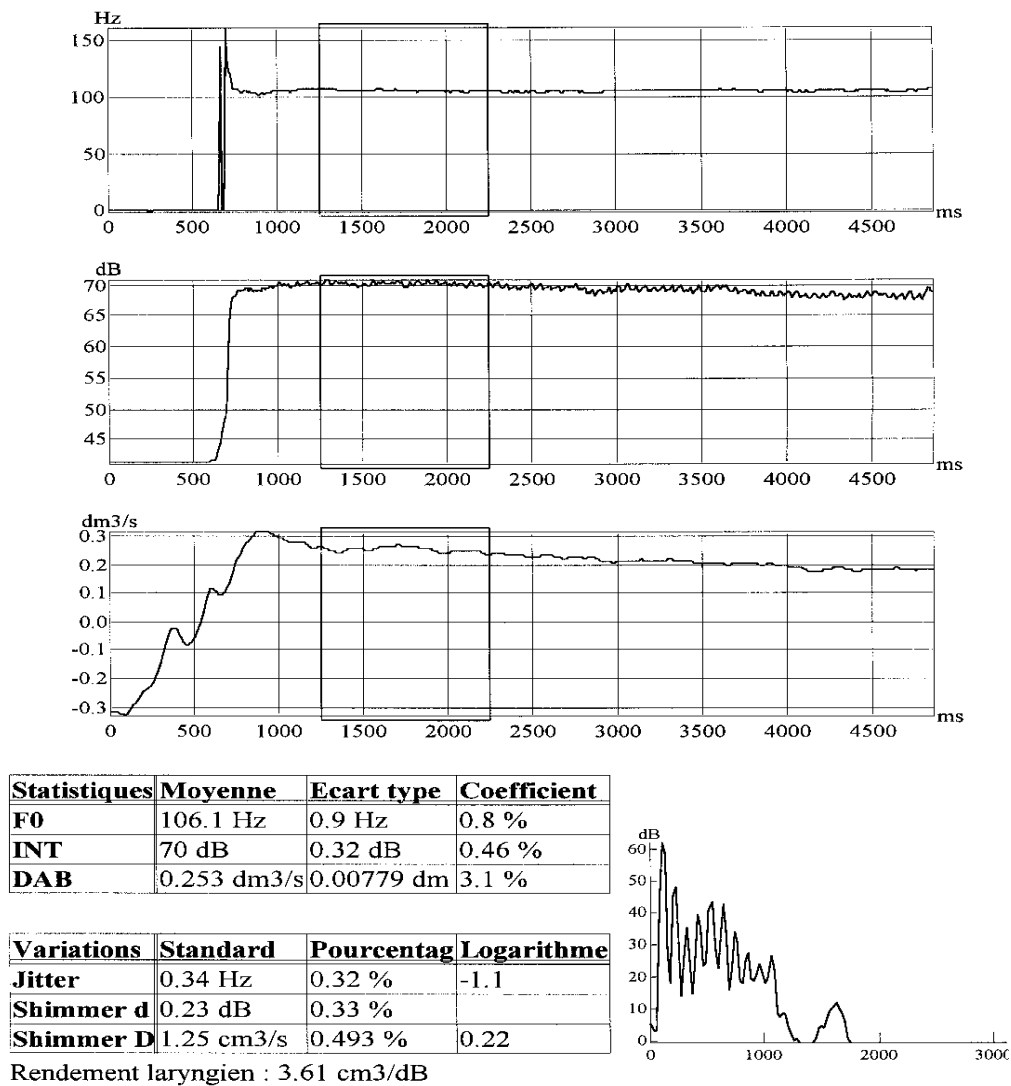
La moyenne de l'intensité de 70 dB montre une voix faible produite « à l'économie », son shimmer de 0.3 % se situe bien en dessous de la frontière de normalité.

En revanche, la moyenne du débit d'air est de 0.253 dm3/s, valeur trop forte même pour une voix d'homme et au-dessus de la frontière pathologique. La fuite glottique de 3,61 cm3/dB se situe aux limites de la normalité, elle démontre un faible rendement laryngien pour un sujet « normal ». On constate sur le spectre une faible émergence harmonique avec une forte composante de bruit. Ce dernier est, contrairement au cas précédent, un bruit de souffle, à cause du fort débit d'air oral moyen et non un bruit d'instabilité vibratoire à cause du faible jitter. Un tel spectre dénote une voix de grade 0 sur l'échelle GRBAS d'évaluation des dysphonies, mais légèrement soufflée à l'écoute.

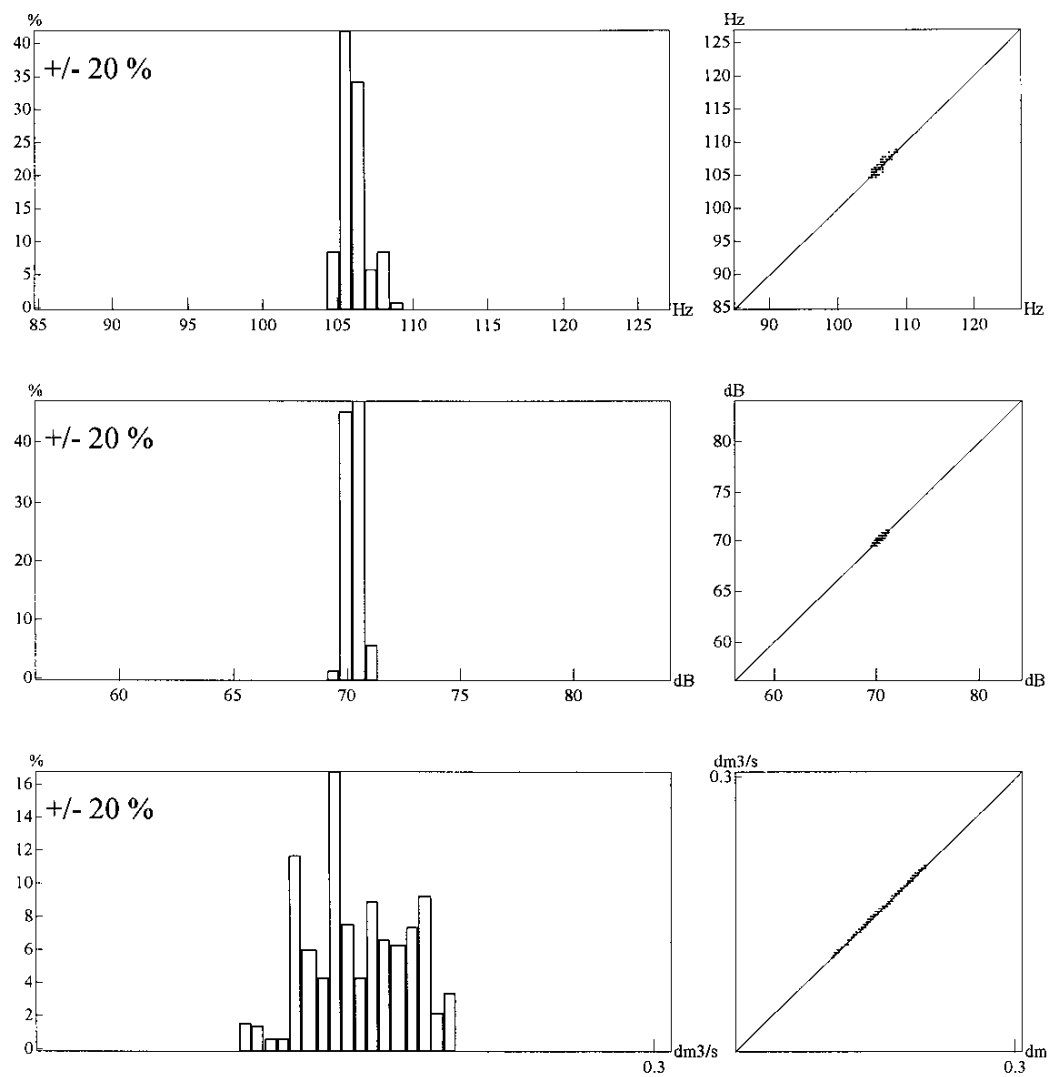
La figure 16-B montre des distributions et portraits de phase particulièrement resserrés sur les trois paramètres. Ceci confirme un fonctionnement stable du vibrateur glottique à comparer avec le cas précédent.

La figure 16-C représente le profil vocal de ce sujet « normal ». On remarque que l'essentiel du dysfonctionnement glottique affecte dans ce cas, au-delà du cercle marquant la frontière de normalité, les axes du débit d'air oral et du bruit provoqué par la fuite glottique.

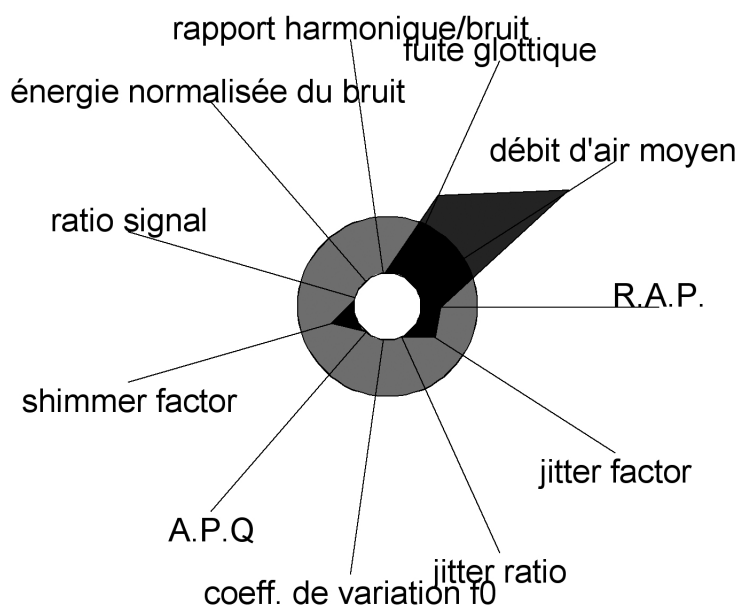
Ce cas démontre que, même considérée comme normale à l'examen perceptif, une voix peut avoir des caractéristiques objectives pathologiques. En fait, ce locuteur « normal » a une voix légèrement voilée sans être dysphonique, cotée 0 sur l'échelle GRBAS. C'est la voix fatiguée d'un enseignant après une journée de cours.



De haut en bas : Les variations de la mélodie ou  $F_0$  en Hertz (Hz), celles de l'intensité de la voix en décibel (dB), et celles du débit d'air oral en centimètre cube par seconde (cm<sup>3</sup>/s). L'axe des abscisses est en millisecondes. Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités. Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1250 et 2250 millisecondes. Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation.



**Figure 16-B**  
*Distributions des variations des paramètres  $F_0$ , intensité et débit d'air oral du cas précédent*



**Figure 16-C**  
*Profil vocal de l'état de dysphonie  
à travers l'amplitude des différents paramètres du cas précédent*

La figure 17-A représente de haut en bas : les variations des paramètres  $F_0$ , intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée de 2.5 secondes par un patient parkinsonien. Son niveau de dysphonie selon l'échelle GRBAS est estimée au grade 1.

L'étude du décours de ces trois paramètres montre, tout d'abord, une très faible durée d'émission vocale. On constate une instabilité de la fréquence fondamentale à moyen terme sous la forme d'un lent tremblement de l'ordre de 3 Hz. L'intensité et surtout le débit ont une déclivité très importante de l'ordre de 4/5 pour ce dernier, en moins d'une seconde après le début du voisement. Ceci démontre une mauvaise coordination pneumophonique. Les mesures statistiques sur ces trois paramètres sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1400 et 2400 millisecondes.

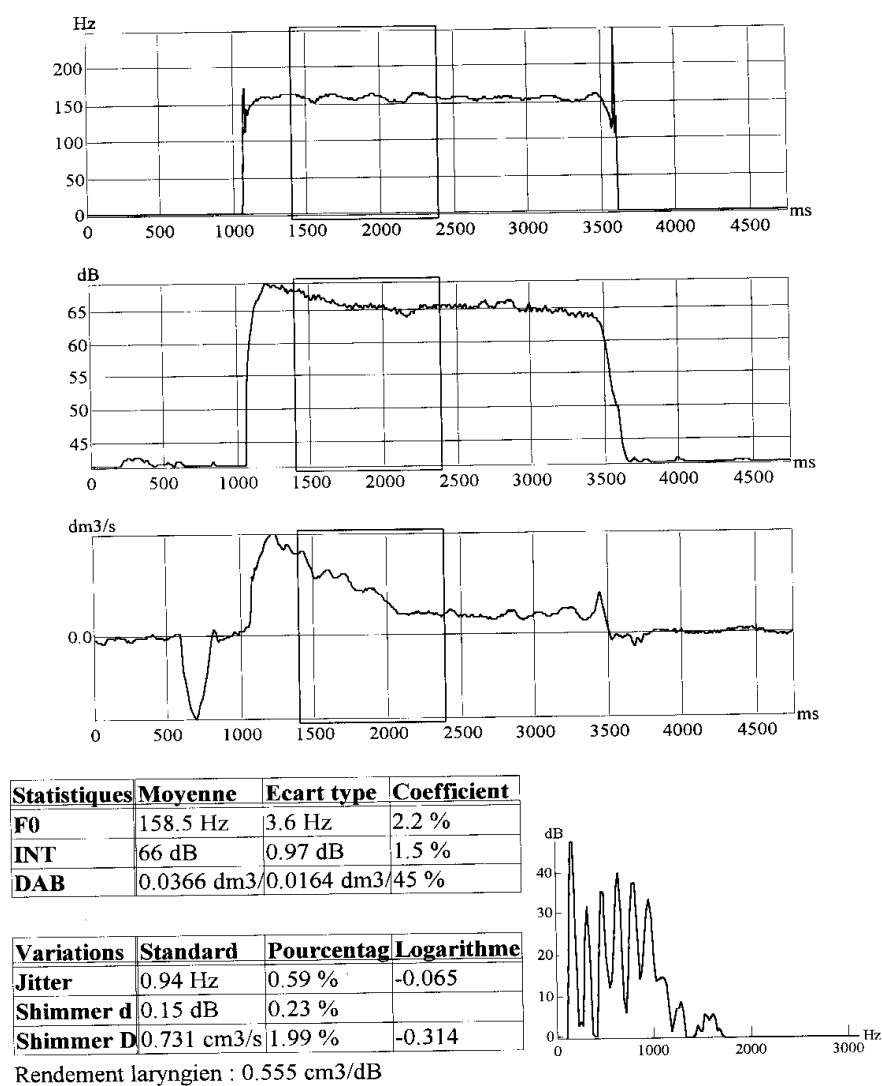
La moyenne de la  $F_0$  de 158 Hz est normale pour une voix d'homme. Son jitter (factor) à 0.59%

est également normal, par contre son écart-type de 3.6 Hz et son coefficient de variation de 2.2 sont élevés et confirment les instabilités à moyen terme, décrites précédemment. La moyenne de l'intensité de 66 dB montre une voix un peu faible, son shimmer de 0.23 % étant normal.

La moyenne du débit d'air de 0.036 dm<sup>3</sup>/s est très faible pour une voix d'homme. On doit noter la valeur faible de la fuite glottique de 0.555 cm<sup>3</sup>/dB, qui démontre un très bon rendement, bien meilleur que dans le cas précédent. Cet exemple montre que certaines voix peu dynamiques et utilisant un faible débit d'air peuvent avoir un rendement laryngien élevé malgré une médiocre audibilité. On constate sur le spectre une faible émergence harmonique et une composante de bruit équivalente au cas précédent. Cependant ici, il ne peut être provoqué ni par une fuite glottique, ni par un fort jitter. Il est dû à la modulation à moyen terme de la Fo (coefficient de variation) et à l'état hypertonique du larynx qui dégrade la qualité du signal glottique. La voix est sourde et voilée.

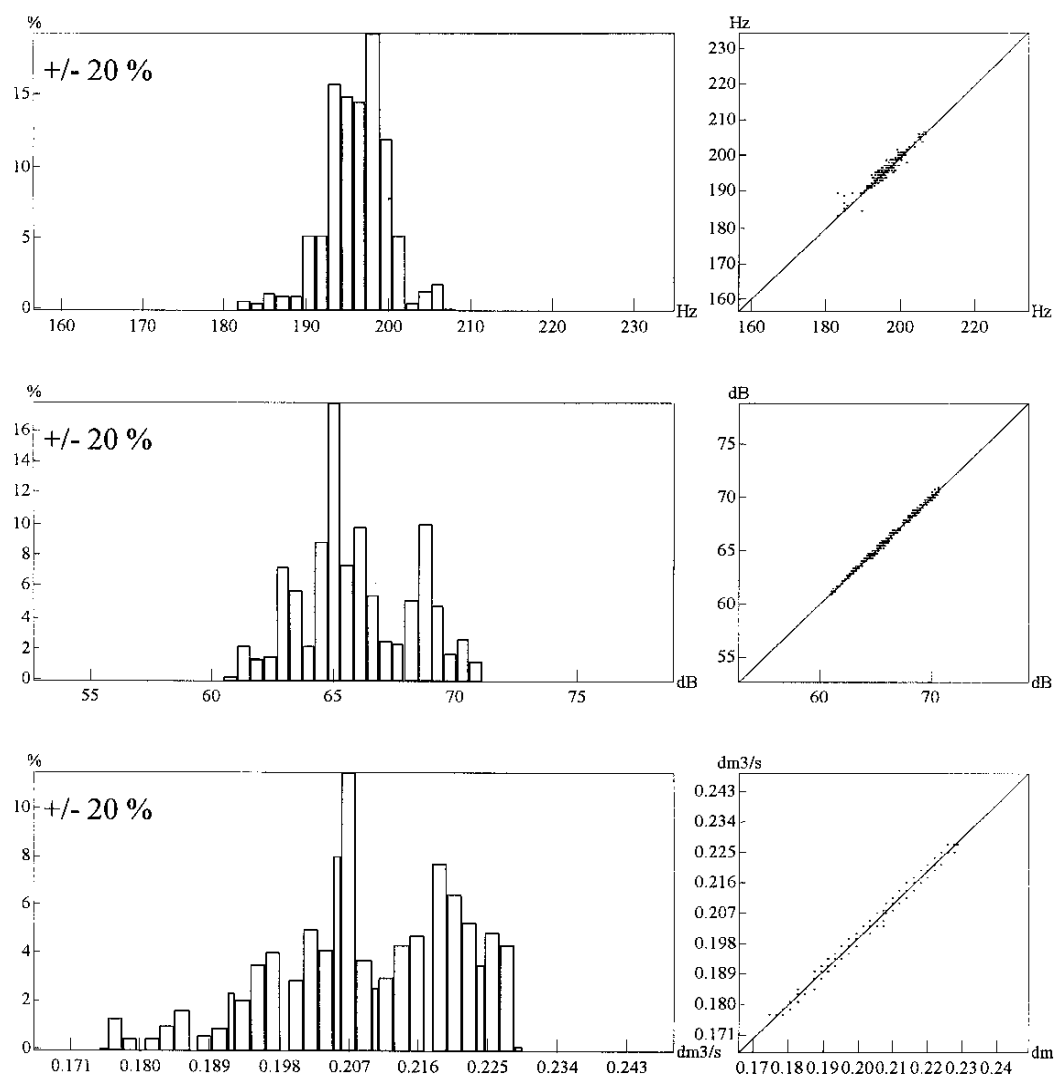
Les distributions de la figure 17-B confirment les remarques précédentes et ne montrent que des variations à long et moyen termes sur les trois paramètres.

On remarque sur le profil vocal de la figure 17-C que l'essentiel du dysfonctionnement glottique de ce patient affecte au-delà du cercle marquant la frontière de normalité les 3 indices de bruit et le coefficient de variation de la Fo.

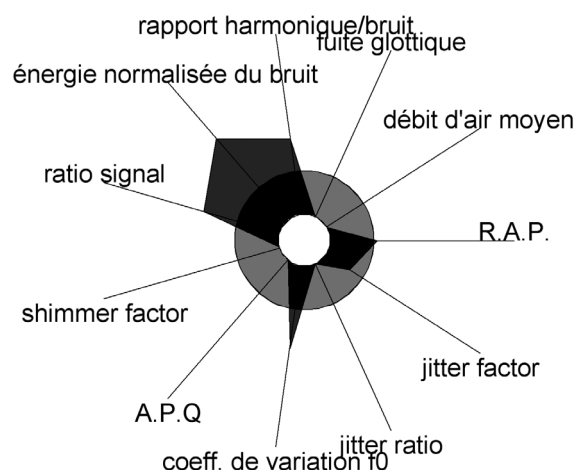


De haut en bas : Les variations de la mélodie ou Fo en Hertz (Hz), celles de l'intensité de la voix en décibel (dB), et celles du débit d'air oral en centimètre cube par seconde (cm³/s). L'axe des abscisses est en millisecondes (ms). Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités. Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 1400 et 2400 millisecondes. Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation.

**Figure 17-A**  
Evaluation vocale d'un patient parkinsonien sur la production d'un « a » tenu



**Figure 17-B**  
*Distributions des variations des paramètres  $F_0$ , intensité et débit d'air oral du cas précédent*



**Figure 17-C**

*Diagramme synthétique de l'état de dysphonie à travers l'amplitude des différents paramètres du cas précédent*

La figure 18-A représente de haut en bas : les variations des paramètres  $F_0$ , intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée de dix secondes par un patient souffrant d'un syndrome parkinsonien, plus précisément de la maladie de Steele-Richardson qui provoque une forte dysphonie de grade 3 selon l'échelle GRBAS. Sa voix est quasiment inintelligible.

L'étude du décours de ces trois paramètres montre une très grande instabilité générale de la production vocale. D'abord, le patient est incapable de tenir la voyelle sans interruption du voisement. Sur une durée inférieure à 10 secondes, on en observe 4. Il contrôle très mal les attaques de voisement pour lesquelles on constate des transitoires d'amplitude et de durée importants. Ces phénomènes sont caractéristiques d'une forte hypertonie laryngée. Ensuite, les parties pseudo stables de l'émission vocale sont affectées par une importante modulation en phase, et synchrones sur les 3 paramètres. La plus grosse affecte le paramètre intensité avec une dynamique de 5 dB environ. Cela laisse supposer que ces modulations ont pour origine un tremblement de la mandibule qui, agissant sur l'aperture aux lèvres, module directement l'intensité et le débit d'air oral, et indirectement la  $F_0$ . En effet, lorsque l'aperture aux lèvres diminue, l'intensité et le débit diminuent également, par contre la pression intra orale augmente, ce qui a pour conséquence de diminuer la  $F_0$ .



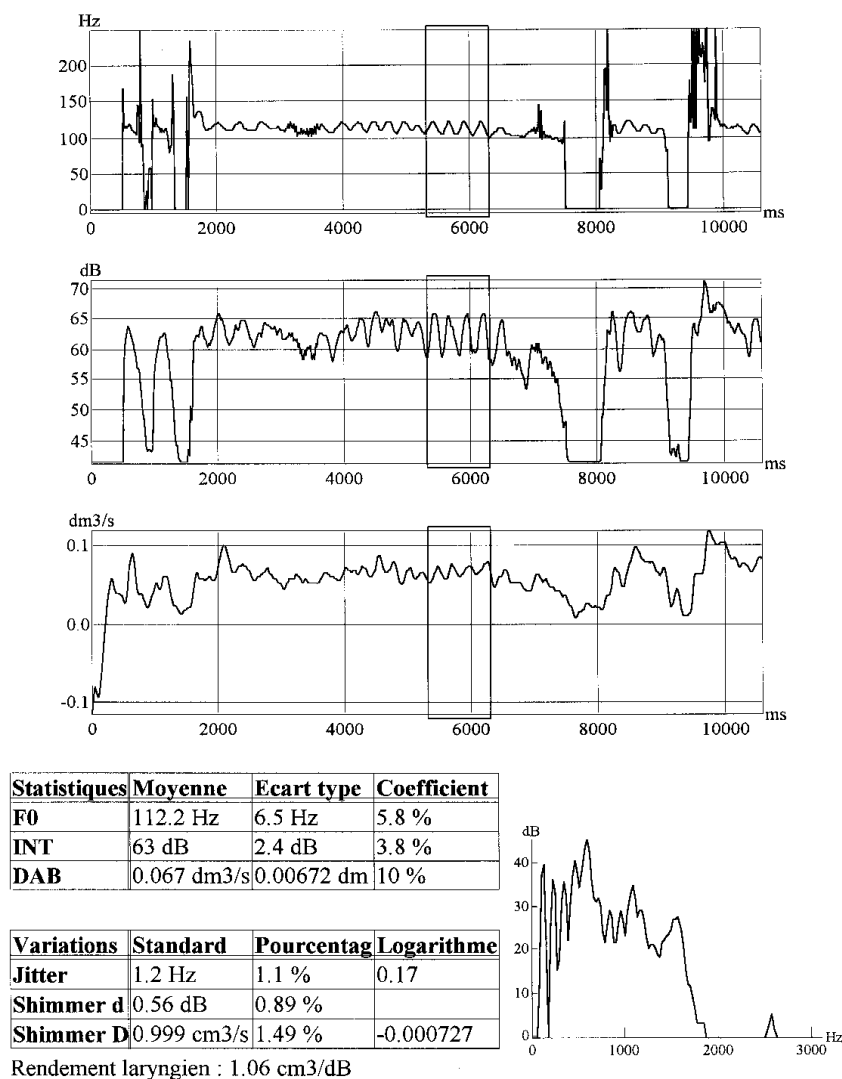
Les statistiques élémentaires sur ces trois paramètres sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 6250 et 7250 millisecondes.

La moyenne de la Fo de 112 Hz est normale pour une voix d'homme, par contre son écart-type de 6,5 Hz, et surtout son coefficient de variation de 5,8%, confirment la grande instabilité à moyen terme de la Fo entrevue à la lecture des courbes. Ces valeurs sont d'ailleurs aux limites supérieures de la pathologie (Tableau 1).

La moyenne de l'intensité de 63 dB montre une voix faible, son shimmer de 0,9% se situe à la frontière de normalité, ainsi que le jitter (factor) à 0,9 %, ce qui témoigne d'une faible instabilité à court terme du vibrateur laryngien.

La moyenne du débit d'air est de 0,067 dm<sup>3</sup>/s, valeur normale pour une voix d'homme. Sa variabilité est également normale pour un débit. Le rendement laryngien (fuite glottique) est de 1,06 cm<sup>3</sup>/dB, ce qui est normal. On constate sur le spectre une faible émergence harmonique avec une forte composante de bruit. Ce dernier est un bruit d'instabilité à cause du fort coefficient de variation de la Fo et non un bruit de souffle, à cause du faible débit d'air oral moyen. On remarque dans la figure 18-B que l'histogramme de la Fo semble montrer une bimodalité très forte de la distribution des vibrations de la glotte comme dans la figure 15-B et peu laisser supposer un mode vibratoire différent pour chaque corde vocale. En fait, il n'en est rien, la distribution « bimodale » est provoquée par la modulation périodique de la Fo. Seule l'étude du portrait de phase peut montrer la différence des deux phénomènes. Dans la figure 15-B, la bimodalité de la Fo se manifeste par deux nuées de part et d'autre de la diagonale alors que dans ce cas, les deux nuées se situent sur la diagonale, preuve qu'elles sont dues à des variations à moyen terme. Ceci démontre la complémentarité de l'histogramme et du portrait de phase. On retrouve également cette pseudo bimodalité sur l'histogramme de l'intensité pour la même raison.

Le profil vocal de la figure 18-C montre que l'essentiel du dysfonctionnement glottique de ce patient affecte, au-delà du cercle marquant la frontière de normalité, les 3 indices de bruit et le coefficient de variation de la Fo, comme dans le cas précédent mais avec une plus grande ampleur.



De haut en bas : Les variations de la mélodie ou F0 en Hertz (Hz), celles de l'intensité de la voix en décibel (dB), et celles du débit d'air oral en centimètre cube par seconde (cm3/s).

L'axe des abscisses est en millisecondes (ms).

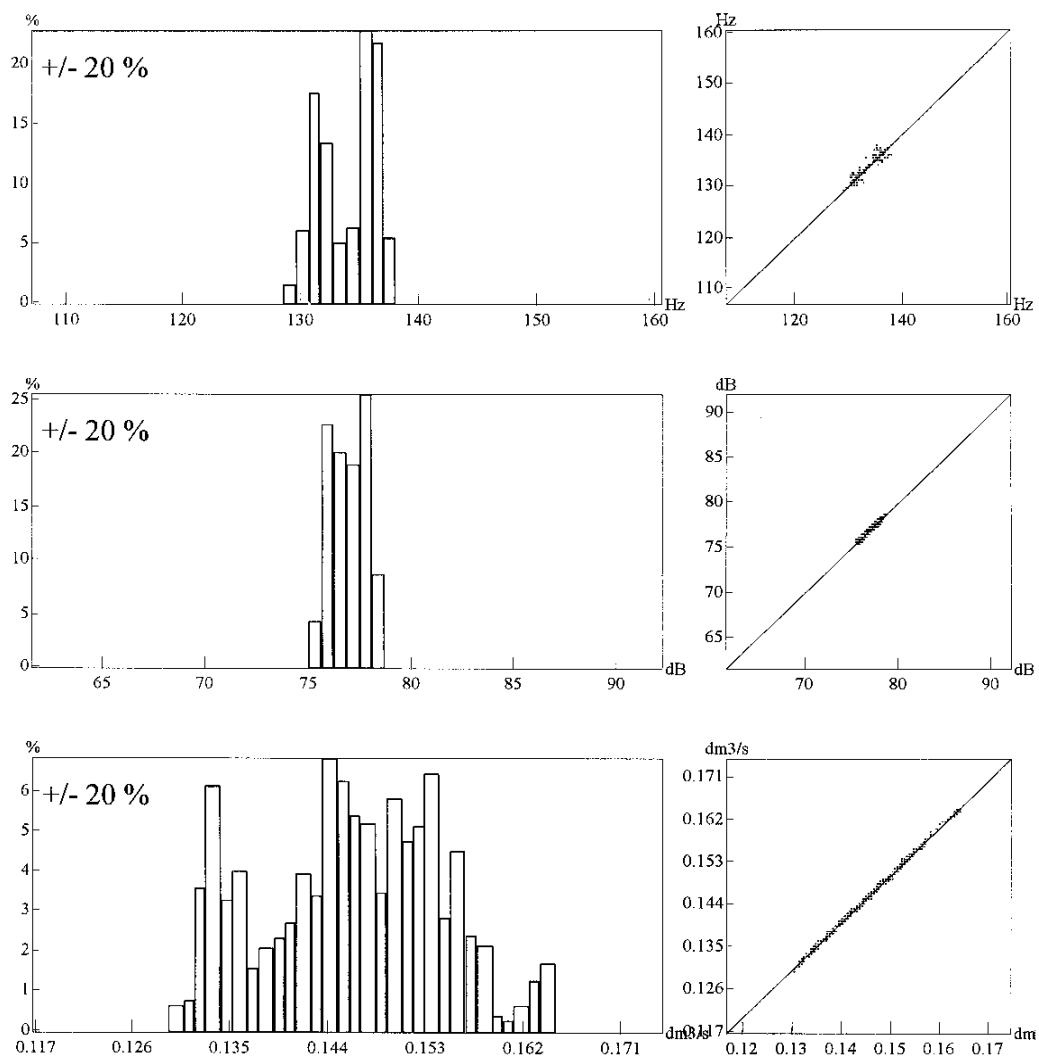
Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités.

Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 5200 et 6200 millisecondes.

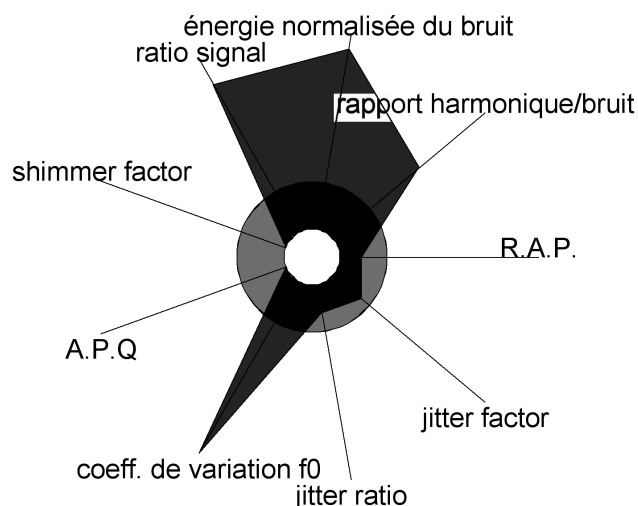
Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation.

**Figure 18-A**

*Evaluation vocale d'un patient souffrant d'un syndrome parkinsonien sur la production d'un « a » tenu*



**Figure 18-B**  
*Distributions des variations des paramètres  $F_0$ , intensité et débit d'air oral du cas précédent*



**Figure 18-C**  
*Profil vocal synthétisant l'état de dysphonie à travers l'amplitude  
des différents paramètres du cas précédent*

La figure 19-A représente les variations des paramètres  $F_0$ , intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée de six secondes par une patiente souffrant de dystonie laryngée. La dysphonie sévère qui en résulte est cotée de grade 3 selon l'échelle GRBAS. Elle est caractérisée par une voix très déstabilisée.

L'étude du décours de ces trois paramètres montre une grande difficulté du larynx de cette patiente à entrer en vibration. Tout d'abord il y a une abondante production d'air oral avant que le larynx ne commence à vibrer de manière erratique avec de nombreux dévoisements. Il atteint une phase de vibration à peu près stable 3 secondes après le début de la production vocale volontaire. Ceci montre une très mauvaise coordination pneumophonique. Cependant, la voix, une fois établie, semble être d'assez bonne qualité. Les mesures sont effectuées sur une partie de cette phase vocale stable d'une seconde de durée dans la fenêtre comprise entre 3000 et 4000 millisecondes.

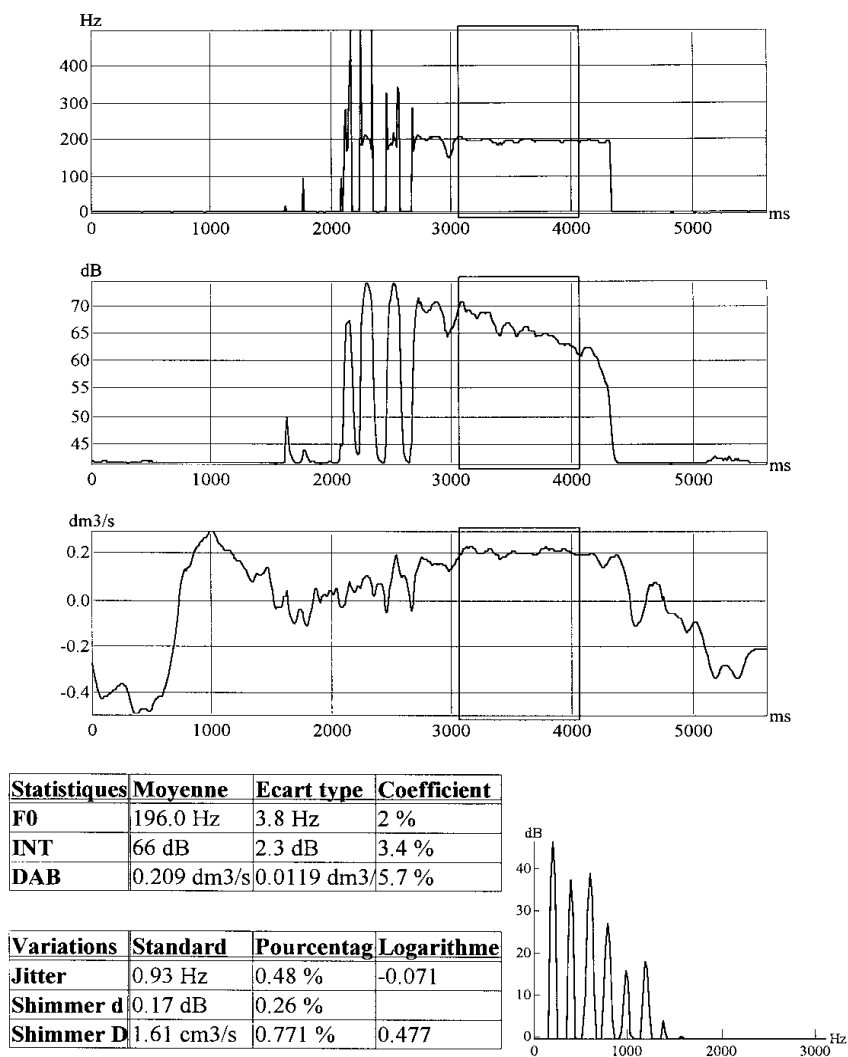
La moyenne de la  $F_0$  de 196 Hz est normale pour une voix de femme ainsi que son jitter (factor) de 0,5 %. Par contre, son écart-type de 3,8 Hz et son coefficient de variation de 2 % témoignent d'une instabilité à moyen terme au dessus de la frontière de pathologie. La moyenne de l'intensité de 66 dB montre une voix un peu faible au shimmer normal de 0,26 %. La moyen-

ne du débit d'air est de 0.209 dm<sup>3</sup>/s, valeur aux limites de la normale pour une voix de femme, ainsi que le rendement laryngien (fuite glottique) de 3.17 cm<sup>3</sup>/s/dB.

Le spectre montre une très bonne émergence harmonique avec une absence presque totale de bruit. Il indique une voix claire et bien timbrée.

La figure 19-B montre des histogrammes et portraits de phase qui confirment des variations à moyen terme sur les 3 paramètres sans autres informations particulières.

Le profil vocal de la figure 19-C indique que l'essentiel du dysfonctionnement glottique de cette patiente affecte au-delà du cercle marquant la frontière de normalité, les indices aérodynamiques du débit d'air oral et de la fuite glottique ainsi que le coefficient de variation de la Fo.



De haut en bas : Les variations de la mélodie ou  $F_0$  en Hertz (Hz), celles de l'intensité de la voix en décibel (dB), et celles du débit d'air oral en centimètre cube par seconde (cm3/s).

L'axe des abscisses est en millisecondes.

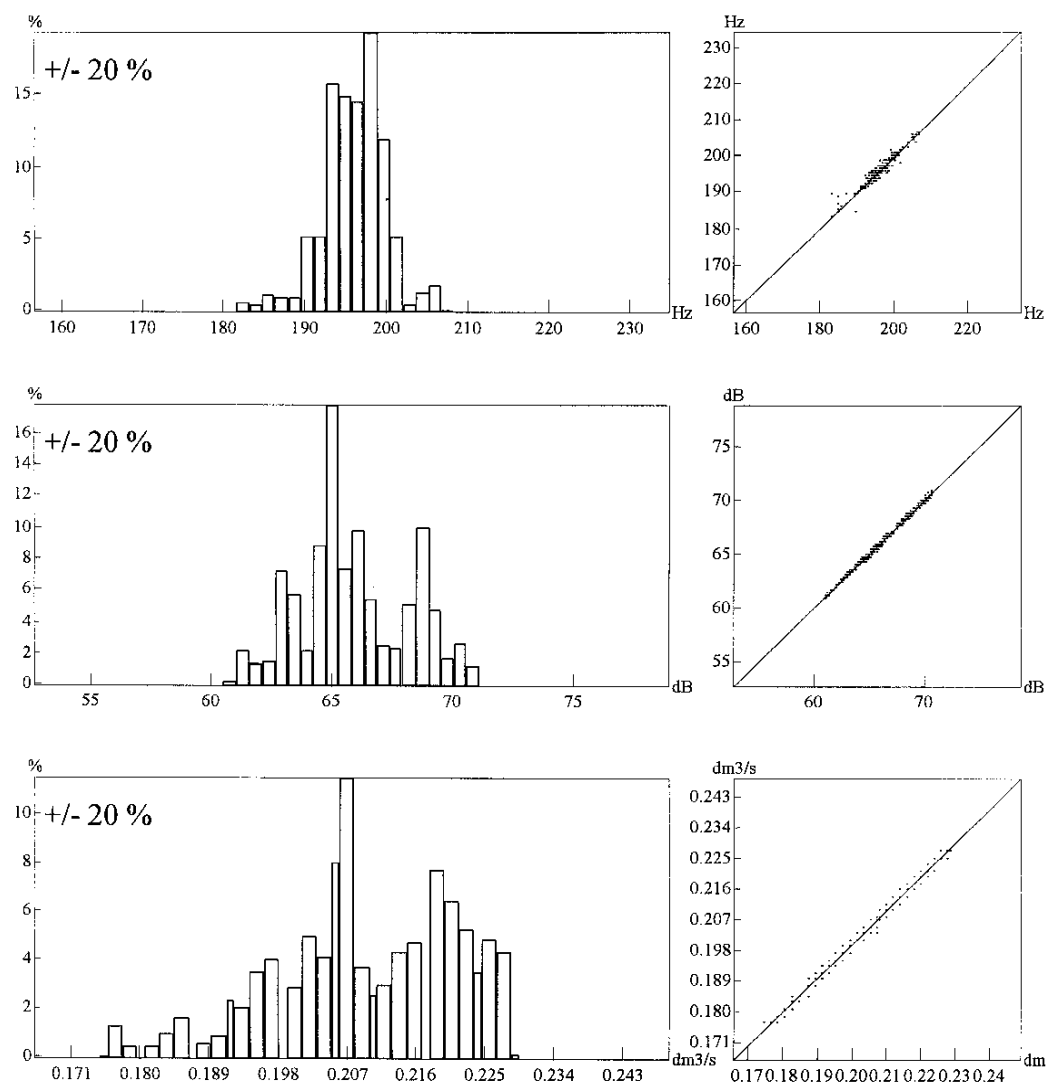
Le premier tableau contient des statistiques élémentaires sur ces trois paramètres, le second contient leurs variabilités.

Ces mesures sont effectuées sur une durée de 1 seconde dans la fenêtre comprise entre 3050 et 4050 millisecondes.

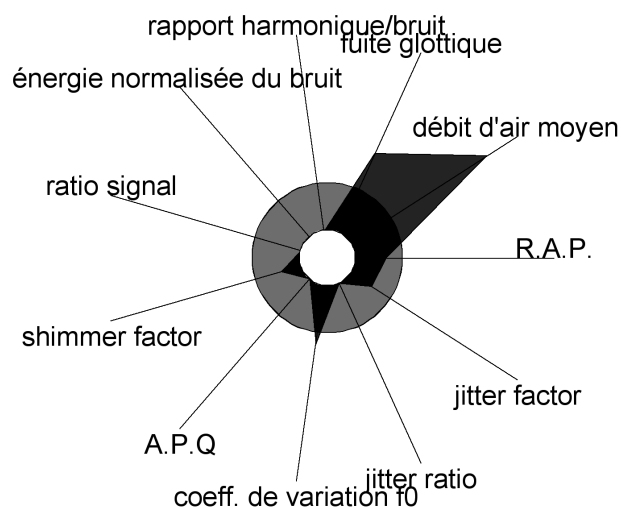
Le spectre en bas à droite est le spectre moyen en bande étroite sur la même fenêtre d'observation.

### Figure 19-A

Évaluation vocale sur une voyelle "a" tenue de l'état de vibration d'un larynx dystonique



**Figure 19-B**  
*Distributions des variations des paramètres  $F_0$ , intensité et débit d'air oral du cas précédent*



**Figure 19-C**  
*Profil vocal synthétisant l'état de dysphonie à travers l'amplitude  
des différents paramètres du cas précédent*

La figure 20 représente un exemple d'évaluation vocale au moyen du système EVA 2. Il s'agit d'une patiente souffrant d'une paralysie de la corde vocale gauche, qui provoque une forte dysphonie de grade 3 selon l'échelle GRBAS, caractérisée par une voix rauque, soufflée, accompagnée de multiples désonorisations.

La figure 20-A représente de haut en bas : les variations des paramètres, signal de parole,  $F_0$ , intensité et débit d'air oral lors de l'émission d'un « a » tenu sur une durée inférieure à trois secondes.

L'étude du décours à long terme de ces quatre paramètres montre une grande difficulté à émettre une voyelle stable. La patiente a fait de tels efforts pour mettre son larynx en action qu'elle a crié, à tel point que le manipulateur a été surpris pour le réglage du niveau d'entrée et que le signal de parole a saturé. La fenêtre d'observation d'une seconde entre 1575 et 2575 millisecondes a été choisie sur une partie non écrêtée du signal acoustique. L'émission vocale a duré moins de deux secondes pour se terminer dans un souffle. Les décours de la  $F_0$  et de l'intensité montrent une atténuation progressive depuis le début du voisement jusqu'à son extinction. Ceci est dû à la chute très rapide du débit d'air oral dont la valeur glisse en moins de 3 secondes de 1 dm<sup>3</sup>/s à zéro. On constate une assez forte instabilité de la fréquence fondamentale à court et moyen terme que confirment sa distribution et son portrait de phase sur la figu-

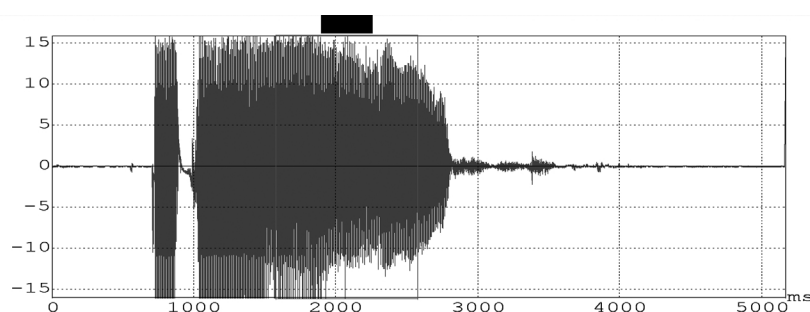


re 20-B. La figure 20-C regroupe les tableaux statistiques des paramètres  $F_0$ , intensité, débit et bruits. On constate que la moyenne de la  $F_0$  de 165 Hz est normale quoique un peu grave pour une voix de femme, par contre, son écart-type de 3.86 Hz, son jitter (factor) de 1.73% et son coefficient de variation de 2.33% confirment la grande variabilité de la  $F_0$  entrevue à la lecture des courbes.

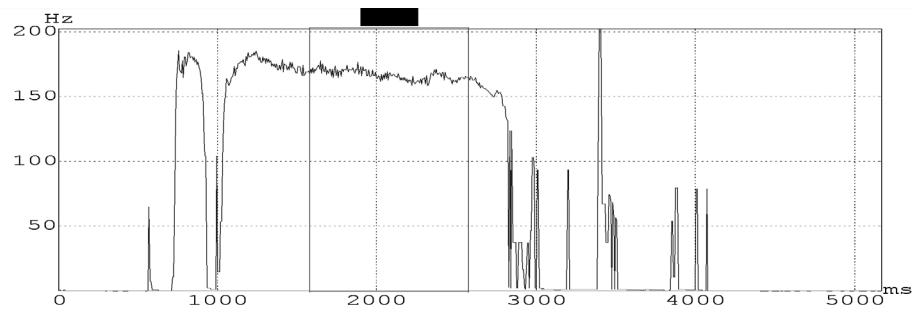
La moyenne de l'intensité de 90 dB montre une voix forte, une voix criée, forcée. Son shimmer semble ne pas avoir été détecté mais il est certainement faible, à la lecture de la courbe d'intensité.

La moyenne du débit d'air est de 0.550 dm<sup>3</sup>/s, valeur très importante au maximum de la pathologie (Tableau 1) qui explique la très courte durée de l'émission vocale. La fuite glottique de 6 cm<sup>3</sup>/s/dB est également très forte. L'étude du spectre moyen de la figure 20-B, en bande étroite sur la même fenêtre d'observation, nous montre une faible émergence harmonique qui ne va pas au-delà de 1 kHz, avec une forte composante de « bruit ». Ce dernier est un bruit d'instabilité à cause du jitter et surtout un bruit de souffle, à cause de son niveau d'énergie constant sur toute la largeur du spectre et le très important débit d'air oral moyen.

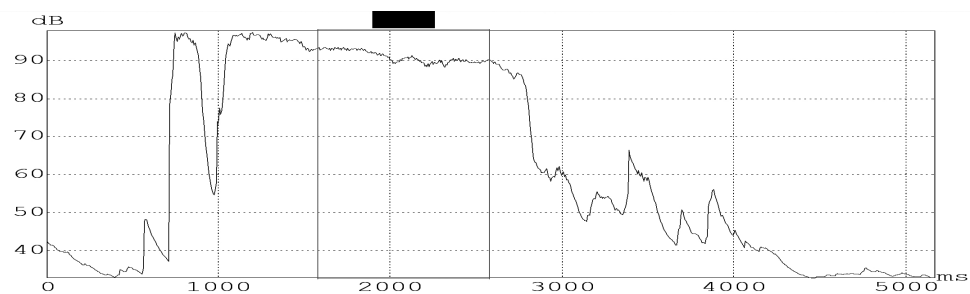
La figure 20-C représente le profil vocal qui synthétise schématiquement tous les indices mesurés à partir des paramètres  $F_0$ , intensité et débit. On remarque que l'essentiel du dysfonctionnement glottique affecte, au-delà du cercle marquant la frontière de normalité, les indices de bruit, d'instabilité vibratoire et surtout de débit.



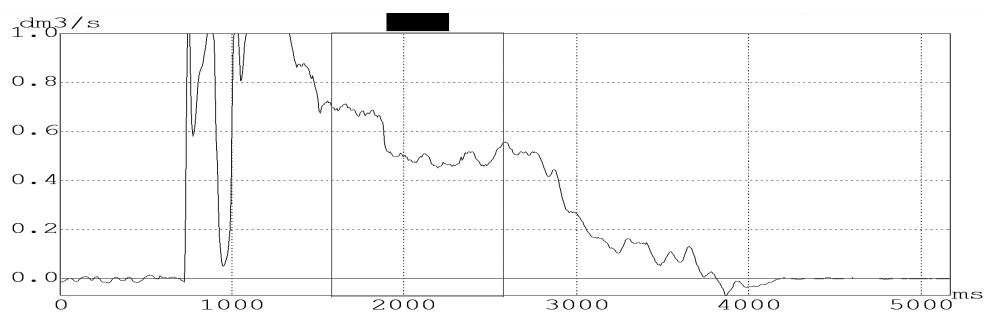
*Variations du signal acoustique (ordonnées sans dimension, abscisse en millisecondes [ms])*



*Variations de la fréquence fondamentale en Herz (Hz)*



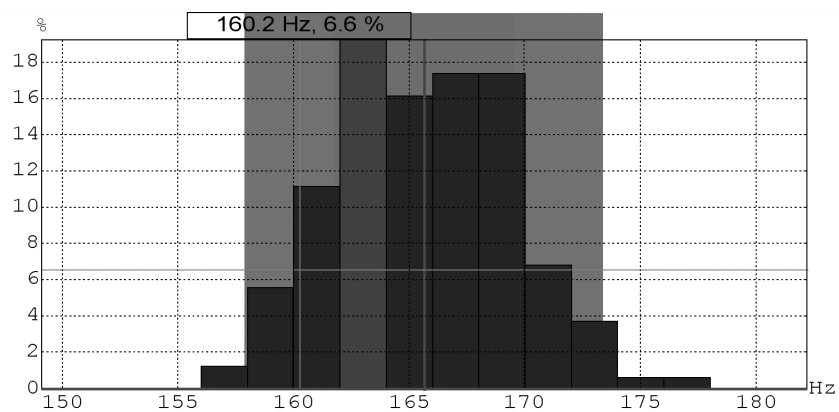
*Variations de l'intensité en décibels (dB)*



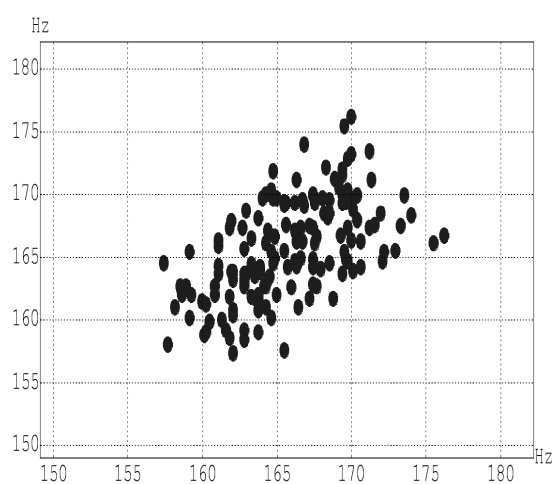
*Variations du débit d'air oral en décimètre cube par seconde (dm³/s)*

### **Figure 20-A**

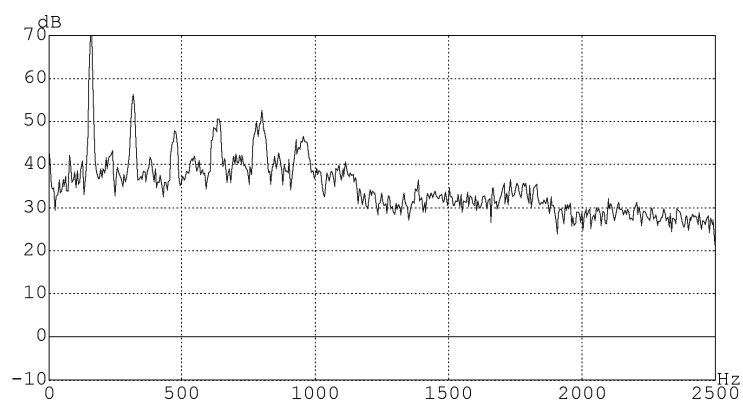
*Evaluation vocale sur une voyelle "a" tenue de l'état de vibration du larynx  
d'une patiente souffrant d'une paralysie cordale*



*Histogramme de la distribution de la fréquence fondamentale*



*Portrait de phase de la fréquence fondamentale.*



*Spectre moyen en bande étroite sur la durée de la fenêtre d'observation.*

**Figure 20-B**

*Distributions des variations du paramètres Fo et spectre du cas précédent*

Fréquence fondamentale	
Fo moyenne	165.6 Hz
Ecart type de la Fo	3.86 Hz
Coefficient de variation de la Fo	2.33 %
Jitter absolu	2.87 Hz
Jitter factor	1.73 %
Jitter ratio	17.25
R.A.P	0.01

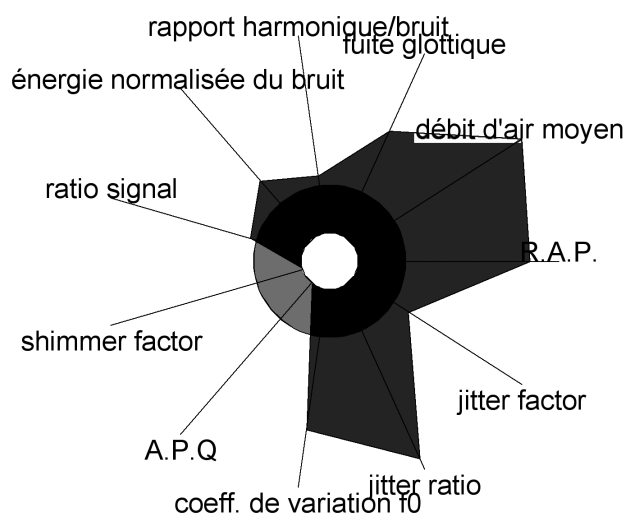
Débit d'air oral	
Débit d'air moyen	0.550 dm <sup>3</sup> /s
Ecart type du débit	0.091 dm <sup>3</sup> /s
Coefficient de variation du débit	16.41 %
Fuite glottique	6.05 cm <sup>3</sup> /s/dB

Intensité	
Intensité moyenne	90.98 dB
Ecart type de l'intensité	1.41 dB
Coefficient de variation de l'intensité	1.55 %
Shimmer absolu	0.3 dB
Shimmer factor	0.8 %
A.P.Q	0.08

Analyse du bruit [1575-2575 ms]	
Rapport harmonique/bruit (H/N ratio)	5.74 dB
Energie normalisée du bruit (NNE)	-4.23 dB
Ratio signal	62.2 %
Ratio signal [f>1 kHz]	5.0 %
Fo détectée	157 Hz

**Figure 20-C**

*Tableaux des différents indices du cas précédent, calculés sur la durée de la fenêtre d'observation, à partir des paramètres Fo, intensité, débit d'air oral et analyse acoustique*



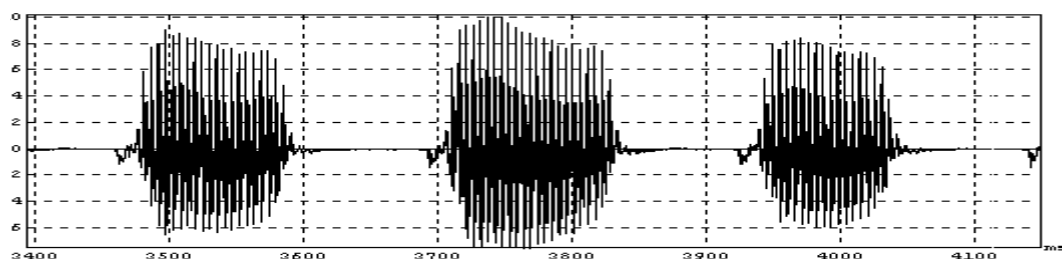
**Figure 20-D**  
*Profil vocal synthétisant l'état de dysphonie à travers l'amplitude  
des différents paramètres du cas précédent*

La figure 21 montre trois exemples d'évaluation de l'efficacité glottique. Elle consiste à faire prononcer au locuteur une suite de syllabes « pa ». Sur la durée d'émission d'une voyelle « a » sont mesurés : le débit d'air oral ainsi que l'intensité moyenne. La valeur de la pression intra orale est mesurée pendant l'occlusion du « p » qui précède la voyelle. Cette valeur peut être considérée comme celle de la pression sous-glottique car la glotte est ouverte pendant l'occlusion labiale du « p ». On en déduit différents indices tels que l'efficacité, le rendement et la résistance glottique qui donnent une bonne évaluation de l'hypertonie (forçage vocal) ou de l'hypotonie du larynx.

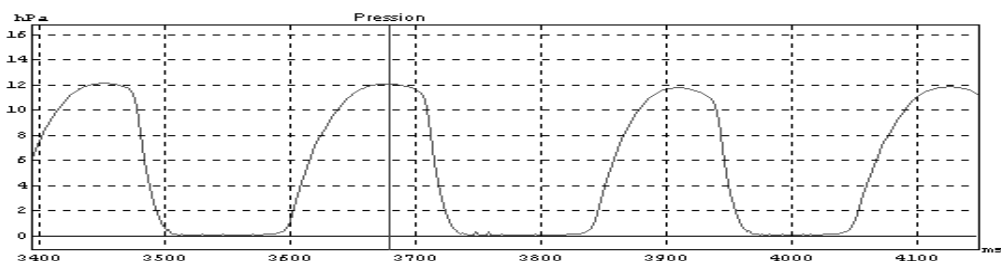
La figure 21-A représente l'évaluation de l'efficacité laryngée du locuteur normal de la figure 16. Chez ce locuteur, on constate une bonne articulation labiale caractérisée par des montées de débit d'air oral sur les explosions de « p » rapides et amples. La pression intra-orale (sous-glottique estimée) de 12.2 hPa est un peu élevée, mais bien corrélée avec l'intensité de l'émission vocale de 99 dB qui indique une voix forte, une voix projetée. Le débit d'air est de 0.320 dm<sup>3</sup>/s. Il est un peu élevé, mais ce sujet a précédemment montré la même particularité, ainsi que pour la valeur de la fuite glottique qui va dans le même sens.

La figure 2I-B représente l'efficacité laryngée du patient parkinsonien de la figure 17. On constate, contrairement au cas précédent, une mauvaise articulation avec une absence d'occlusion labiale pendant l'ouverture de la glotte, et des explosions de « p » très peu marquées réduites à un bruit de souffle. Ceci explique une très faible valeur de la pression intra-orale de 3,5 hPa associée à une intensité moyenne de 76 dB et un faible débit d'air de 0,04 dm<sup>3</sup>/s. Cependant, lorsque le patient réalise bien l'occlusion labiale comme dans le « p » à l'extrême droite de la figure, la valeur de la PIO ne monte pas au-dessus de 5,5 hPa. Cette faible PIO, associée à un très faible débit oral, donne un rendement glottique extravagant de 497, valeur très élevée en rapport avec un produit PIO x Débit très faible. Ces indices dénotent une forte hypotonie.

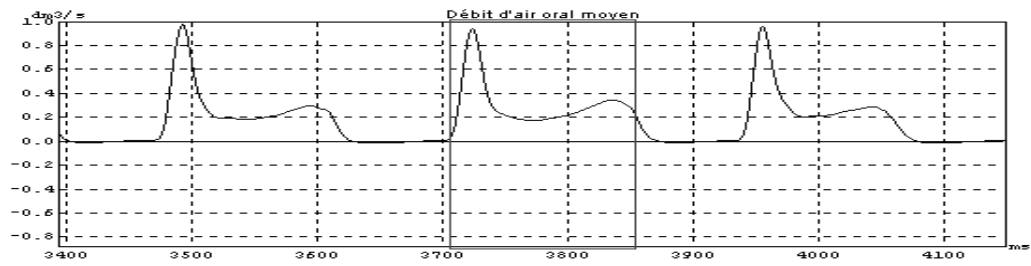
La figure 2I-C représente l'efficacité glottique d'un patient ataxique à la suite d'un accident vasculaire cérébral. On constate une bonne occlusion labiale avec une très ample valeur du débit d'air oral, surtout sur la voyelle. La valeur de la pression intra orale de 20 hPa est également très forte, associée à une intensité moyenne normale de 83 dB. Par opposition, le rendement glottique est très faible. Ce tableau témoigne d'une hypertonie importante.



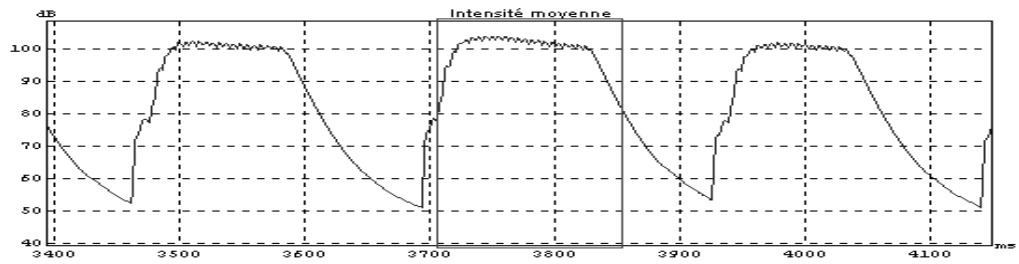
*Signal acoustique d'une répétition de « **pa, pa, pa, pa,...** »  
L'axe des abscisses est en millisecondes.*



*Courbe de variation de la pression intra orale (l'axe des ordonnées est en hecto Pascal [hPa])*



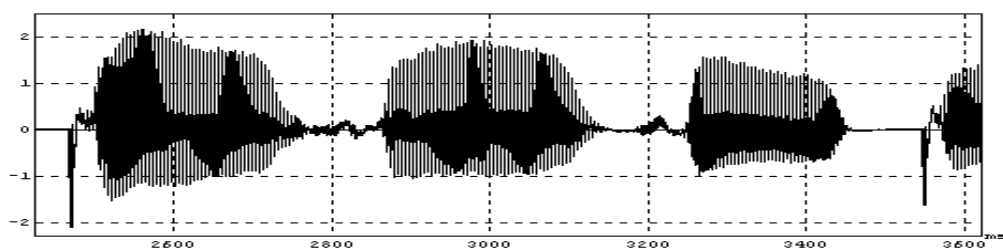
*Courbe de variation du débit d'air oral*  
*L'axe des ordonnées est en décimètre cube par seconde (dm³/s).*



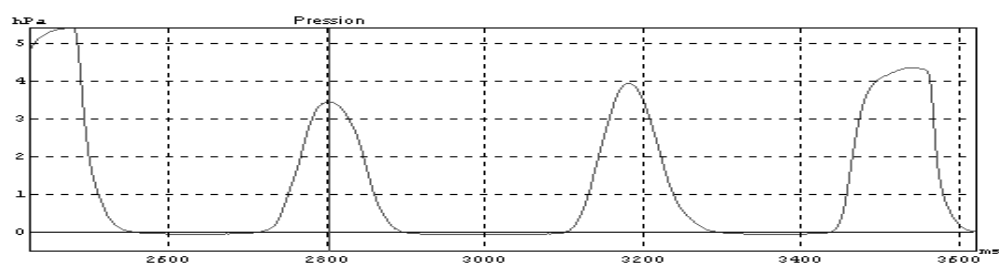
*Courbe de variation de l'intensité moyenne (RMS) du signal acoustique.*  
*L'axe des ordonnées est en décibel (dB).*

	Pression intra orale	Intensité moyenne	Débit d'air oral moyen	Fuite glottique	Efficacité glottique	Rendement glottique	Résistance glottique
Unité	hPa	dB	dm³/s	cm³/dB	dB/hPa	dB/ (hPa.dm³/s)	hPa/ (dm³/s)
Valeur	12.2	99.4	0.32	3.21	8.18	25.59	38.12
Temps	3679 ms	3705+[147] ms	3705+[147] ms				

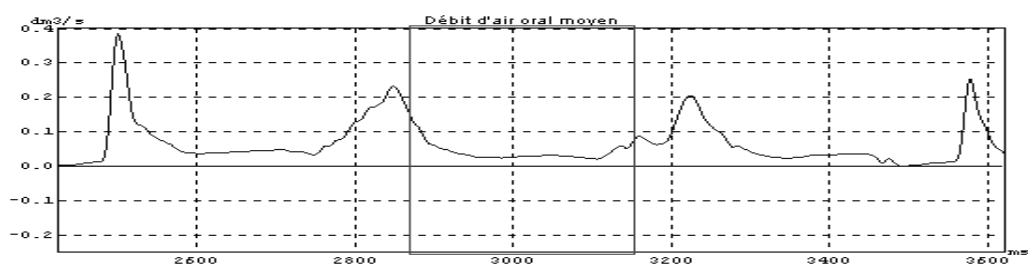
**Figure 21-A**  
*Evaluation de l'efficacité laryngée d'un sujet normal*



*Signal acoustique d'une répétition de « **pa, pa, pa, pa,...** »  
L'axe des abscisses est en millisecondes.*

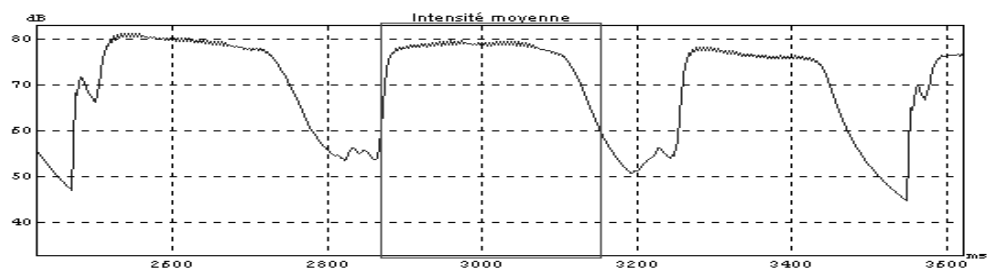


*Courbe de variation de la pression intra orale (l'axe des ordonnées est en hecto Pascal [hPa])*



*Courbe de variation du débit d'air oral  
L'axe des ordonnées est en décimètre cube par seconde ( $dm^3/s$ ).*

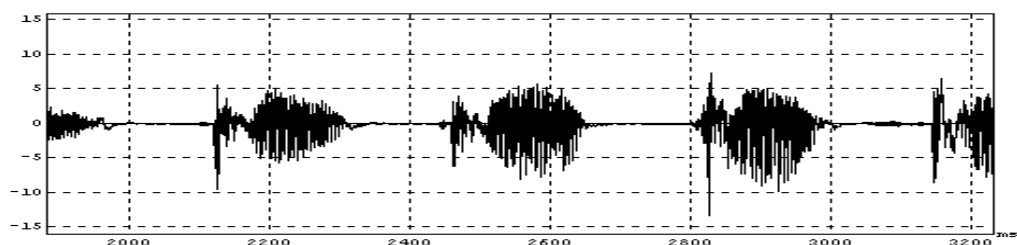




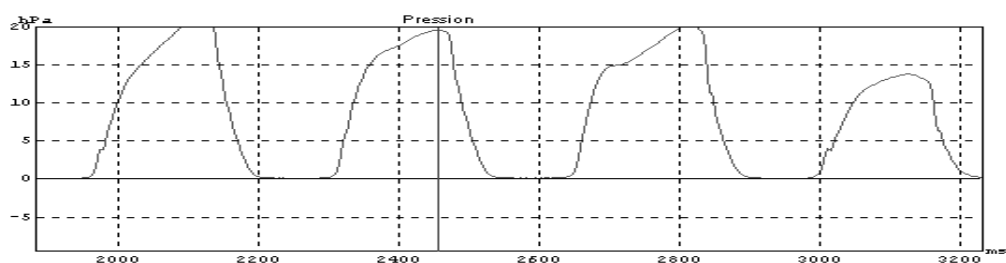
*Courbe de variation de l'intensité moyenne (RMS) du signal acoustique  
L'axe des ordonnées est en décibel (dB).*

	Pression intra orale	Intensité moyenne	Débit d'air oral moyen	Fuite glottique	Efficacité glottique	Rendement glottique	Résistance glottique
Unité	hPa	dB	dm <sup>3</sup> /s	cm <sup>3</sup> /dB	dB/hPa	dB/ (hPa.dm <sup>3</sup> /s)	hPa/ (dm <sup>3</sup> /s)
Valeur	3.5	75.8	0.04	0.53	22.18	497.51	75
Temps	2803 ms	2857+[-285] ms	2857+[-285] ms				

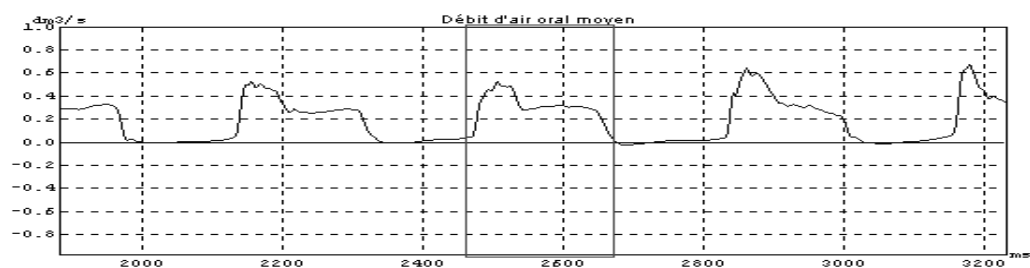
**Figure 21-B**  
*Evaluation de l'efficacité laryngée d'un patient parkinsonien*



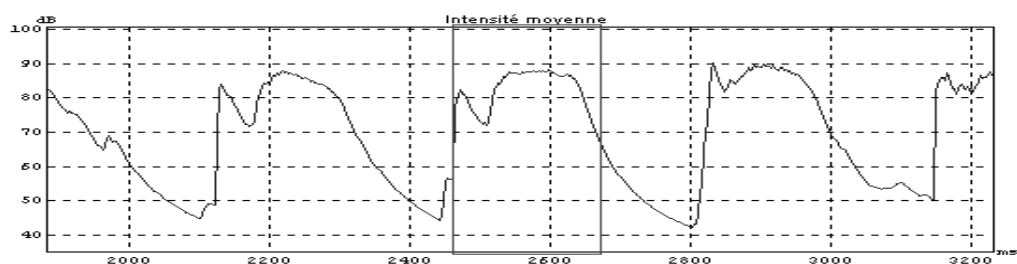
*Signal acoustique d'une répétition de « **pa, pa, pa, pa,...** »  
L'axe des abscisses est en millisecondes.*



*Courbe de variation de la pression intra orale (l'axe des ordonnées est en hecto Pascal [hPa])*



*Courbe de variation du débit d'air oral*  
*L'axe des ordonnées est en décimètre cube par seconde (dm³/s).*



*Courbe de variation de l'intensité moyenne (RMS) du signal acoustique.*  
*L'axe des ordonnées est en décibel (dB).*

	Pression intra orale	Intensité moyenne	Débit d'air oral moyen	Fuite glottique	Efficacité glottique	Rendement glottique	Résistance glottique
Unité	hPa	dB	dm³/s	cm³/dB	dB/hPa	dB/ (hPa.dm³/s)	hPa/ (dm³/s)
Valeur	19.6	82.5	0.32	3.85	4.20	13.33	61.25
Temps	2454 ms	2460+[209] ms	2460+[209] ms				

**Figure 21-C**  
*Evaluation de l'efficacité laryngée d'un patient ataxique (suite d'AVC)*

#### 4.3.7. Les analyses multiparamétriques

Les atteintes pathologiques pouvant avoir une influence sur la qualité de la voix affectent simultanément plusieurs de ses dimensions physiques. Ces différentes dimensions sont affectées de manière très variable en fonction de leurs origines physiopathologiques. Elles contiennent donc des informations complémentaires sur les atteintes pathologiques. Pour mieux évaluer leur importance et améliorer la mise en relation des mesures objectives avec les évalua-

tions perceptives de jurys d'experts, les phoniâtres ont naturellement cherché à associer plusieurs mesures objectives complémentaires, pour en déduire des indices plus discriminants que des paramètres particuliers. Ces études ont porté surtout sur l'association de paramètres acoustiques et aérodynamiques (Dejonckere, 1990 ; Giovanni *et al.*, 1996 ; Piccirello *et al.*, 1998). Ces auteurs concluent tous à l'importance de l'information apportée par les paramètres aérodynamiques dans l'évaluation des dysphonies. Certaines études portent sur l'association de paramètres uniquement acoustiques pour en définir un indice de sévérité de la dysphonie (Wuyts *et al.*, 2000). C'est dans la direction des analyses multiparamétriques que les évaluations objectives des dysphonies progressent. Le temps n'est plus où certains essayaient de promouvoir un paramètre particulier, persuadés que l'essentiel de l'information du dysfonctionnement se trouvait dans une unique dimension du signal vocal. Les exemples que nous avons commentés précédemment montrent bien la complémentarité et tout l'intérêt des évaluations multiparamétriques.

Dans sa thèse de Sciences, Yu Ping (2001) a établi que 81 % des patients avaient une évaluation objective (à partir du système EVA) en accord avec un jury d'experts chargé d'évaluer l'état pathologique de leur voix. Nous atteignons là les limites théoriques des méthodes d'évaluations perceptives. Elle en conclut qu'il est désormais possible et fondé de recommander aux phoniâtres l'utilisation de méthodes d'évaluation objectives des dysphonies.

## Conclusion

Malgré leurs imperfections, personne ne peut contester les progrès dont ont bénéficié les évaluations objectives des dysarthries et des dysphonies ces dix dernières années. Les méthodes acoustiques ont gagné en efficacité et en simplicité d'utilisation et d'interprétation. Les méthodes aérodynamiques ont également fait de nombreux progrès, en particulier au plan de la mise en œuvre et de la fiabilité. Ces progrès sont dus bien évidemment à ceux de la micro-informatique mais principalement à trois autres raisons.

D'une part, les cliniciens se sont habitués aux informations objectives. Ils se sont familiarisés avec leurs interprétations et surtout les ont fait évoluer dans le sens d'une meilleure analyse physiopathologique. Enfin, ils commencent à avoir confiance dans les mesures objectives, en leur réservant une place raisonnable dans leurs protocoles d'évaluation clinique.

D'autre part, les acousticiens et linguistes, à qui l'on doit la grande majorité des méthodes de mesures utilisées dans les évaluations objectives, ont de leur côté fait des efforts particuliers pour adapter leurs outils de recherche à des utilisations cliniques en collaborant de manière plus approfondie avec les praticiens.

Pour finir, quelques entreprises ont investi beaucoup d'efforts dans des instruments d'évaluation spécifiques, homologués pour des utilisations médicales et capables d'effectuer des mesures fiables en utilisation clinique de routine grâce à une ergonomie adaptée.

Cependant, les évaluations objectives peuvent et doivent encore s'améliorer dans trois directions. D'abord dans la réalisation de bases de données homogènes pour en déduire des bases de connaissances indispensables à une meilleure interprétation des dysfonctionnements articulaires et vocaux. Ensuite, par une mise en relation plus poussée entre les mesures effectuées et les phénomènes physiopathologiques. Enfin, par la définition de standards d'évaluation plus précis et consensuels. La tâche à réaliser est encore importante mais elle est bien engagée et progresse inéluctablement.

## 5. L'enregistrement numérique de la voix et la parole : problèmes et méthodes

Etudier un son quel qu'il soit, en l'écoutant attentivement, consiste d'une manière générale à porter particulièrement son attention sur le timbre (c'est-à-dire son spectre de fréquence), sur la précision des attaques sonores, sur la modulation de sa mélodie, et sur sa dynamique (c'est-à-dire le rapport entre sa plus petite et sa plus grande amplitude). Pour permettre l'écoute à posteriori autant de fois qu'il est nécessaire pour saisir toutes les subtilités des sons et surtout, la comparaison de séquences sonores émises à des moments différents, de multiples dispositifs d'enregistrements sonores ont été imaginés, dont la technique magnétique.

### 5.1. L'enregistrement magnétique

Il consiste à transformer le signal acoustique capté par un microphone en un champ magnétique, analogue au signal initial, qui s'inscrit sur une bande magnétique défilant à une certaine vitesse. Cette trace du signal demeure sur la bande tant qu'un nouveau champ magnétique ne l'efface pas. Il est possible ensuite de transformer la trace magnétique enregistrée sur la bande en un signal acoustique en la faisant défiler à nouveau, à la même vitesse, et par l'intermédiaire d'un haut parleur. Malheureusement l'enregistrement magnétique n'a pas permis pendant de longues années une restitution satisfaisante du signal original pour nos applications.

Le timbre d'un son enregistré est fonction de la distorsion harmonique de l'enregistreur, c'est-à-dire de sa propension à créer dans le signal des harmoniques qui n'existent pas. On s'accorde à considérer comme satisfaisante une distorsion harmonique inférieure à 0.1 %. Le timbre est également, ainsi que les attaques sonores, dépendant de la largeur de bande en fréquence

et de sa linéarité. On s'accorde généralement sur des valeurs de 10 kHz pour la parole et de 20 kHz pour la musique à + ou - 1,5 dB.

La stabilité de la mélodie est subordonnée à un défilement de la bande magnétique précis et stable, dont le "pleurage" doit être inférieur à 1%. Enfin, la dynamique est tributaire surtout du bruit de fond (qui est, pour sa plus grande partie, fonction de la qualité du support magnétique). La dynamique d'un son est définie en décibels (dB), elle est de l'ordre de 80 dB pour la parole, 100 dB pour le chant et 120 dB pour la musique symphonique.

Depuis à peine une vingtaine d'années, l'enregistrement sonore magnétique a permis de conserver le signal de parole de manière fiable, commode et pour un prix "domestique" au moyen de magnétophones à cassettes. Auparavant, la qualité des enregistrements magnétiques nécessitait des matériels onéreux, lourds et peu pratiques, qui étaient l'apanage d'institutions professionnelles ou d'amateurs fortunés.

Cependant, malgré son prix abordable, l'enregistrement magnétique sur cassette "institutionnel" (grâce surtout aux progrès des bandes magnétiques dont il existe des standards différents) n'a pas toutes les qualités, en particulier pour des études comparatives de voix et de parole normale ou pathologique. Si la largeur de bande, la linéarité en fréquence et la stabilité de défilement sont devenues acceptables, la distorsion harmonique (2%), le bruit de fond restent trop importants et la dynamique (55 dB) trop faible pour saisir à l'écoute toutes les subtilités des constituants du signal vocal. Evaluer à l'écoute le souffle d'une voix lorsque la bande magnétique en a encore plus, n'est pas très efficace. Enfin, l'enregistrement sur bande vieillit mal. Avec le temps, le signal prend du bruit et de l'écho. Il existe des techniques de traitement du signal pour augmenter artificiellement la dynamique (DOLBY par exemple). Elles ont une réelle efficacité pour la musique domestique mais sont à éviter dans nos applications car elles apportent des distorsions supplémentaires non contrôlées.

## 5.2. L'enregistrement numérique (ou digital)

La technique de l'enregistrement numérique constitue la grande famille de l'audio numérique avec le CD Audio (Compact Disk laser Audio), le DAT (Digital Audio Tape) et le MiniDisc. Elle a été la grande révolution de l'enregistrement audio au cours des années 80. Cette technique d'enregistrement numérique s'apparente à l'enregistrement sur ordinateur dont elle utilise les mêmes principes et bénéficie des mêmes qualités qui permettent de dépasser les normes développées précédemment. Nous avons mentionné, à propos de l'enregistrement magnétique, que le signal enregistré sur la bande était analogue au signal acoustique initial. Il s'agit là d'une technique d'enregistrement dite analogique. La technique numérique (ou digi-

tale) consiste à transformer le signal acoustique en une suite de nombres dont la valeur seule est enregistrée. Cette technique ne rajoute pas de bruit au signal et ne le distord quasiment pas. Elle a ainsi presque doublé la dynamique, augmenté la bande passante en fréquence avec une linéarité parfaite et diminué la distorsion harmonique et le pleurage à des niveaux quasi non mesurables. La technique numérique nécessite des traitements particuliers qui sont l'échantillonnage et la quantification. L'échantillonnage consiste à prendre la valeur du signal acoustique capté par un microphone, à intervalles réguliers. Cette opération se fait à une certaine fréquence dont la valeur fixe la largeur de bande en fréquence du signal. En effet, il ne peut pas exister dans un signal de composantes supérieures à la moitié de la valeur de la fréquence d'échantillonnage. La quantification consiste à transformer l'amplitude de chaque échantillon en un nombre codé en numération binaire c'est-à-dire avec un certain nombre de bits. Cette opération limite la dynamique du signal. En effet cette dernière est de 6 dB par bit. On appelle conversion analogique-numérique du signal l'ensemble de ces deux opérations.

Par exemple, la norme du CD audio est de 44 kHz pour la fréquence d'échantillonnage et de 16 bits pour la résolution de quantification. Le signal a donc une largeur de bande en fréquence de 22 kHz théorique, mais en pratique de 20 kHz et une dynamique théorique de 96 dB mais en pratique toujours plus faible.

L'enregistrement sur DAT, apparu voici une quinzaine d'années, utilise une bande magnétique comme support sous la forme de mini cassette qui autorise une durée d'enregistrement de 90 minutes en stéréo (2 canaux). Il existe différents types d'enregistreurs professionnels, « de table », portable, ultra portable etc.. mais son principe de tambour tournant, emprunté aux magnétoscopes, ne se prête guère à la miniaturisation, si bien que les modèles portables s'avèrent fragiles et peu fiables. Il est toujours plus onéreux que l'enregistrement magnétique analogique classique, mais ce surcoût est compensé largement par le gain de qualité.

L'enregistrement sur MiniDisk est plus récent. Il a été proposé dès 1992 par Sony mais n'a été diffusé par d'autres constructeurs qu'en 1996. Il n'est utilisé dans nos applications que depuis peu de temps. Son principe s'apparente à celui des disques d'ordinateurs magnéto-optiques réinscriptibles, associé à un système de codage original (ATRAC) qui permet de comprimer dans un rapport 5 l'information contenue dans le disque. Ce support est plus petit qu'un CD ROM, il mesure 6 centimètres de diamètre. Il est contenu dans une boîte d'un encombrement inférieur à celui d'une disquette d'ordinateur. Il permet une durée d'enregistrement de 74 minutes en stéréo et le double en monoral (ce que ne fait pas le DAT). Ce support est particulièrement robuste car il est bien protégé dans sa cassette. À son apparition, le système de compression du MiniDisk a été la cause d'une certaine méfiance de la part des utilisateurs pro-

fessionnels. Pour l'avoir testé sur de petits appareils portables de type « baladeur », nous pouvons affirmer qu'il n'apporte aucune distorsion rédhitoire dans nos applications. Là encore, il existe plusieurs types d'enregistreurs ; professionnels, « de table » et portables. Ces derniers sont les plus diffusés sous la forme de baladeurs très performants mais un peu fragiles, qui doivent être manipulés avec soin. Ils sont maintenant à notre avis les instruments idéaux pour les enquêtes linguistiques.

### 5.3. L'enregistrement numérique sur ordinateur

C'est une technique qui existe déjà depuis très longtemps, cependant elle n'a pas toujours été de bonne qualité et elle était surtout réalisée pour effectuer sur le signal de parole des traitements particuliers tels que l'analyse spectrale et la mesure de l'intensité et la mélodie vocale en recherche essentiellement. Ces traitements étaient souvent longs, peu aisés et portaient sur des durées de signaux relativement courtes sauf à utiliser des ordinateurs puissants mais très coûteux. Ce n'est que depuis quelques années que les micro-ordinateurs sont assez puissants pour permettre des enregistrements aux normes de l'audio numérique et des traitements rapides sur les signaux sonores à un prix abordable. Outre la puissance de calcul, l'augmentation des mémoires de masse en particulier permet d'utiliser un micro-ordinateur comme un magnétophone de grande qualité. Les avantages de l'enregistrement sur ordinateurs sont multiples. D'abord, comme il se réduit à un banal fichier informatique, ses propriétés ne varient pas dans le temps ; on peut considérer un tel enregistrement comme indestructible à condition d'être vigilant à sa sauvegarde. Ensuite, il peut être écouté et dupliqué autant que nécessaire sans aucune altération. Enfin, on peut lui appliquer une infinité de traitements, pour réaliser sur les signaux des mesures acoustiques de spectre ou de prosodie par exemple, et des calculs statistiques pour faire sortir du signal des informations très riches imperceptibles à l'écoute. Il est possible également de gérer une quantité considérable d'enregistrements sonores pour les trier, les classer, les comparer ou les mettre en relation avec des informations textuelles et des images sous la forme de bases de données.

Tout ceci n'est cependant possible qu'avec certaines précautions et l'application de principes élémentaires mais fondamentaux.

Le premier de ces principes est d'avoir comme dans le cas d'une chaîne « HIFI » une grande homogénéité de qualité entre les différents constituants du système d'enregistrement numérique.

Ce dernier est centré autour d'un micro-ordinateur de puissance moyenne actuelle (Pentium P3, 1,4 GHz, 128 Mo de RAM, disque 20 Go, Ecran 17", WINDOWS 98 ou 2000) équipé

d'une carte audio à laquelle est connecté un microphone pour enregistrer, et un casque pour écouter. Il est possible de lui adjoindre un graveur de CD ROM pour archiver des enregistrements. Ce support permet, pour un coût de l'ordre d'un euro, d'enregistrer 600 Mo de fichiers audio, c'est-à-dire 12 minutes de signal monophonique au standard CD audio (16 bits, 44kHz de fréquence d'échantillonnage). Contrairement à ce que l'on peut penser, c'est l'ordinateur qui pose le moins de problèmes "acoustiques" et qui constitue l'élément de référence.

#### 5.4. Les fichiers audionumériques

Les fichiers audionumériques sont les fichiers de sons, enregistrés et stockés dans un ordinateur. Il existe une grande variété de formats de fichiers audionumériques pour des raisons historiques, datant d'une époque où chaque concepteur de programme de traitement de sons définissait ses propres fichiers. Au fil du temps, seuls quelques formats de fichiers ont survécu sous la forme de standards tel que le format Wave pour l'audio sur PC et le format AIFF sur APPLE Mac. Cependant, il existe de nombreux formats dérivés surtout des fichiers Wave, dont le plus utilisé est le Wave-PCM. Parfois, les fichiers Wave ne sont pas compatibles entre eux. Il ne faut donc pas s'étonner si d'aventure, on ne peut pas lire des fichiers *.wav* quand on change de matériel d'enregistrement. Seule la connaissance de l'entête du fichier pourra préciser son format exact. La taille des fichiers audionumériques est variable pour une durée d'enregistrement donnée. Elle est fonction de la fréquence d'échantillonnage, du nombre d'octets de codage des échantillons et du nombre de canaux utilisés (mono ou stéréo). La fréquence d'échantillonnage peut être variable. Pour des études sur la parole, elle peut être choisie à 20 kHz, ce qui offre une largeur de bande en fréquence de 10 kHz suffisante pour cette application. Le codage des échantillons doit par contre être réalisé sur 16 bits c'est-à-dire deux octets, pour avoir une bonne définition de la dynamique vocale. Seuls des enregistrements de conversation dénués de tout intérêt acoustique peuvent être codés sur un seul octet. À titre d'exemple, voici quelques tailles de fichiers audionumériques. Le fichier d'un signal monophonique d'une minute, échantillonné à 44 kHz et codé sur 16 bits (comme pour les CD) occupe un espace mémoire un peu supérieur à 5 Mo. Le même signal échantillonné à 22 kHz en occupera deux fois moins et s'il est en stéréo, deux fois plus.

Un format de fichier est à proscrire totalement dans nos applications, il s'agit du format MP3. En effet, ce standard de fichier audionumérique a été développé pour gagner le maximum de place en mémoire pour des signaux de musique, en les comprimant autant que possible dans la limite d'un son pas trop dégradé à l'écoute. Ce qui est acceptable pour de la musique domestique ne l'est pas du tout pour des signaux de parole dont on veut conserver l'intégralité de l'in-



formation. De plus, les fichiers au format MP3 ont des taux de compression qui diffèrent en fonction des systèmes de codage et même s'il est toujours possible de les convertir pour les écouter, ils peuvent être très dégradés et l'on ne peut jamais leur redonner leur qualité initiale.

### 5.5. Les cartes " audio "

On entend sous ce vocable des cartes périphériques amovibles qui permettent au micro-ordinateur de type PC de pouvoir enregistrer et restituer deux signaux sonores (stéréo). Elles sont équipées généralement d'une entrée microphone, d'une entrée et d'une sortie ligne. Cette dernière peut supporter un casque d'écoute mais pas un haut parleur. Elles ont également un connecteur de manettes de jeux et de liaison " MIDI " pour piloter des instruments de musique électronique. Il faut avoir conscience que seules les caractéristiques des entrées-sorties audio nous importent dans nos applications et des caractéristiques " MIDI " extraordinaires ne nous sont d'aucune utilité.

Ces cartes, qui sont qualifiées généralement de " compatibles Soundblaster " du nom du plus ancien et plus diffusé produit standard, représentent l'élément essentiel de la chaîne acoustique. Il ne faut pas s'imaginer que si le son numérique a théoriquement 96 dB de dynamique, 22 kHz de largeur de bande en fréquence et 0.001 % de distorsion harmonique, il en sera de même avec toutes les cartes " son ". Leurs caractéristiques sont très disparates et seuls quelques modèles haut de gamme approchent les précédentes caractéristiques. Certains modèles économiques n'atteignent même pas les caractéristiques d'un bon magnétophone à cassette (Tableau 3).

De plus en plus de micro-ordinateurs sont équipés d'entrées-sorties audio intégrées par construction. Malheureusement elles sont le plus souvent de qualité médiocre, bien que cela soit en train d'évoluer sous la pression de la clientèle toujours plus exigeante. Tous les PC portables sont dans ce cas, non par économie mais par nécessité de miniaturisation. Très rares sont les portables qui disposent d'une acoustique de bonne qualité, même les modèles haut de gamme, à une exception près, la gamme DELL reconnue pour le bon rapport signal sur bruit de ses entrées-sorties audio. Un autre inconvénient des portables réside dans le fait qu'ils sont souvent dépourvus d'entrées " lignes " et cela est très important car il est vivement conseillé de ne pas utiliser directement les entrées microphones qui sont toujours bruyantes. En effet, un micro-ordinateur est un véritable générateur de bruits " audio ". On a beaucoup moins de bruit de fond en amplifiant le signal d'un microphone au moyen d'un préamplificateur extérieur. Si un PC est équipé d'entrée-sortie audio de qualité médiocre, la solution optimale consiste à l'équiper avec une carte audio de bonne qualité. On trouve facilement des cartes de

ce type d'un prix abordable de quelques dizaines d'euros telles que la TURTLE BEACH " Montego ". Dans tous les cas, il faut exiger du distributeur les caractéristiques techniques de la carte son qui équipe votre PC.

## 5.6. Les microphones

Ce sont les capteurs qui transforment les vibrations acoustiques en signaux électriques enregistrables. De leur qualité dépend en grande partie celle du son enregistré. On attend d'un bon microphone d'abord une grande fidélité (linéarité) pour qu'il capte le son avec le moins de distorsion possible. On lui demande ensuite une grande sensibilité associée à un faible bruit car de cela découle une bonne dynamique. Enfin, on exige de lui une bonne robustesse mécanique car c'est l'élément de la chaîne acoustique le plus manipulé.

Le microphone tel que nous l'avons défini existe depuis une vingtaine d'années ; c'est le microphone à électret. Très linéaire, d'une grande dynamique de faible dimension et robuste, c'est le microphone idéal pour notre application. Il est en fait universellement utilisé. Un seul inconvénient cependant, d'ordre pratique : il est nécessaire de l'alimenter par une tension électrique soit par une pile soit par une alimentation " fantôme " extérieure pour assurer son fonctionnement. Celle-ci peut être fournie par un préamplificateur extérieur déjà mentionné, car aucune carte " son " n'en est équipée. Les microphones électrodynamiques, même de qualité, sont à rejeter car ils colorent trop la voix (par leur distorsion) et manquent de dynamique (bruit de fond). Tout microphone économique livré parfois avec des PC est à prohiber impérativement, car ils sont de mauvaise qualité, même ceux livrés avec les logiciels de dictées vocales. Si la dictée fonctionne bien, ce n'est pas un gage de qualité du microphone car les logiciels de reconnaissance " s'habituent " aux microphones autant qu'aux locuteurs.

Il est souhaitable de choisir un microphone qui ait une direction de capture du son privilégiée (microphone directif, unidirectionnel ou cardioïde). En l'orientant vers la source sonore, il est possible d'atténuer les bruits parasites dans l'environnement du locuteur. La place du microphone est également critique. Il est à éviter d'abord de le disposer dans une salle trop réverbérante et près d'une source de bruit, telle que le ventilateur du PC ou une fenêtre, ensuite sur un support tel qu'une table qui transmet très bien tous les bruits (pendulette etc..). Il est nécessaire également de le maintenir à une distance à peu près constante du sujet. Pour cela, ne pas demander au locuteur de tenir lui-même le microphone mais fixer ce dernier sur un pied étudié pour cet usage. Une bonne solution est d'utiliser un micro " cravate ", mais prendre garde alors aux bruits de friction du microphone sur les vêtements qui peuvent être provoqués par les mouvements du locuteur.

La meilleure solution qui intègre tous les avantages est un micro-casque. Ce dispositif très léger est constitué d'un microphone directif fixé par l'intermédiaire d'une liaison ajustable à un support posé sur les oreilles du locuteur. De ce fait, le microphone est toujours à une distance constante de sa bouche et quels que soient les mouvements de sa tête. On le dispose latéralement à quelques centimètres de la commissure des lèvres pour éviter qu'il ne soit sur la trajectoire du débit d'air oral (cela crée des bruits de " flop "). Cette disposition qui privilégie énormément le signal oral atténue par voie de conséquence tous les bruits parasites de l'environnement. Il est possible de calibrer un micro-casque comme un autre microphone en ramenant sa référence à la distance normalisée de 30 centimètres pour faire des mesures sonométriques en décibels absolus et non relatifs. Il en existe plusieurs sur le marché. Leur prix n'est pas supérieur à celui d'un bon microphone. Nous préconisons cependant l'AKG C410 qui existe en version " fantôme " sans alimentation (la moins chère mais qui nécessite un préamplificateur disposant d'une telle alimentation), ou en version alimentée par une pile.

### 5.7. L'écoute

Comme nous l'avons mentionné précédemment, l'écoute du signal enregistré se fait par l'intermédiaire de sorties " ligne ". De nombreuses cartes audio étaient équipées dans le passé de sorties haut parleurs. Ce n'est presque plus le cas maintenant. Nous n'avons pas à le regretter, tellement ces sorties étaient médiocres. Les sorties lignes permettent une écoute normale avec un casque d'impédance moyenne mais manquent souvent de dynamique de restitution, ce qui pour la voix n'est pas très grave. Pour une écoute sur haut parleur il est nécessaire d'utiliser un amplificateur de puissance, une chaîne HI-FI par exemple. On trouve dans le commerce de petits amplificateurs associés à de petites enceintes qui peuvent avoir une bonne qualité. Il faut faire attention aux modèles bas de gamme qui ont une forte distorsion et une largeur de bande en fréquence réduite.

Il existe également quelques dispositifs audio intégrés, externes au PC, qui contiennent des préamplificateurs de microphones avec parfois leurs alimentations " fantômes " et des amplificateurs de sortie ainsi que leur contrôles de gain respectifs. Ces dispositifs se connectent directement aux entrées sorties ligne des cartes audio. Ils représentent la solution la plus efficace au double plan pratique et de la qualité acoustique. Cependant, comme pour tout le reste, il en est d'excellents et de moins bons et l'on doit exiger des fournisseurs leurs caractéristiques techniques. Certains de ces systèmes sont même équipés d'une source sonore de calibrage de microphone.

## 5.8. Les logiciels d'enregistrement-restitution et traitements acoustiques

Les logiciels sont des programmes informatiques qui permettent de réaliser et gérer des tâches ou des applications sur les ordinateurs. La plus fondamentale parmi celles de notre domaine est l'enregistrement et la restitution sonore. Toutes les cartes audio sur PC ont leurs entrées-sorties gérées par le système WINDOWS au moyen des logiciels multimédia natifs. Ils ne sont ni pratiques ni efficaces, par le fait essentiel qu'ils ne permettent pas de contrôler correctement le niveau d'entrée des signaux. Chaque carte audio est livrée avec un ou plusieurs programmes d'enregistrement et un luxe de gadgets pour mixer et trafiquer en tous sens les signaux d'entrée (pour satisfaire les applications audio domestiques axées sur la pop music et la musique synthétique). Là encore il est impossible de contrôler les niveaux d'entrée, les distorsions de tonalité et le mélange des deux voies. L'idéal pour notre application est de disposer d'un réglage de niveau efficace associé à un bon Vu-mètre de contrôle indépendamment sur chaque canal et c'est tout, comme sur un bon vieux magnétophone. Cela n'est même pas le cas de tous les logiciels spécialisés dans les traitements et analyses acoustiques, dont on pourrait attendre un soin particulier au fondamental contrôle des niveaux d'enregistrement. En effet, la saturation ou écrêtage du signal est une " méchante " distorsion irréversible. C'est la plus grave que l'on puisse faire subir à un signal et toute notre attention doit tendre à l'éviter. Un signal écrêté est généralement inutilisable. Il en est de même d'ailleurs pour un signal enregistré trop faiblement. Sans être distordu, il est souvent noyé dans du bruit et on peut très difficilement en obtenir les informations que l'on peut normalement en attendre (analyse spectrale, détection de la Fo etc.). Avant tout archivage d'un enregistrement, ou toute opération de mesure sur ses signaux, il est donc fondamental de bien les cadrer dans la dynamique d'acquisition des entrées des cartes audio.

Nous venons de mentionner pour la première fois le terme de " mesures " et c'est pour les effectuer que toute l'importance de l'ordinateur va se révéler. En effet, il est possible maintenant de réaliser avec un micro-ordinateur des opérations de mesure sur des signaux de parole qui auraient, dans un passé encore proche, nécessité différents instruments spécialisés d'ergonomie disparate et très onéreux. Nous pouvons mentionner entre autres, différents calculs de spectres de fréquences, d'intensité, de durée, de mélodie et de leurs paramètres de stabilité, auxquels peuvent être associés un grand nombre de traitements statistiques pour exprimer leur variabilité. À cela on doit également mentionner les possibilités de classements et de tris automatiques des enregistrements et des résultats des mesures, sous la forme de bases de données. C'est le rôle de programmes particuliers appelés éditeurs de signaux.

Il existe de nombreux éditeurs de signaux susceptibles d'intéresser notre domaine. Nous pou-

vons les classer en trois catégories. D'abord les éditeurs de signaux les plus nombreux visent essentiellement le marché de l'audio numérique domestique ou professionnel. Le produit type en est COOL EDIT. Ces logiciels sont surtout axés sur le montage et sur le traitement esthétique sonore de la musique comme SOUND-FORGE. Ils sont mal adaptés aux mesures de notre domaine même si certaines de leurs fonctions peuvent s'avérer pratiques et s'ils sont généralement ergonomiques et faciles à utiliser. Ensuite, nous trouvons les éditeurs de signaux que nous qualifions de "linguistiques", pour des applications à visées pédagogiques et pour des recherches acoustiques sur la parole en phonétique. WINPITCH en est le produit le plus représentatif de cette famille à côté du grand classique CSL ou MULTI-SPEECH de Kay-Elementric ou encore TFR de Avaaz. Moins nombreux que les premiers, ils ont pour avantage d'être adaptés par définition à l'analyse de la parole. D'un prix variable, mais qui peut être modique, tel que SNORI ou WINPITCH, ils se réduisent cependant aux analyses acoustiques et prosodiques générales et manquent souvent d'ergonomie comme PRATT, produit diffusé gratuitement mais pas maintenu. PHONEDIT est un éditeur de signal un peu particulier par le fait qu'il est le seul à traiter des signaux aérodynamiques et l'électropalatographie. La dernière catégorie des éditeurs de signaux, mais la plus intéressante pour notre domaine, est représentée par les logiciels que nous qualifions de cliniques. On peut, malgré leur petit nombre, définir deux sous-classes. La première est constituée par des éditeurs de signaux généraux auxquels on a greffé des analyses spécifiques au domaine clinique comme MDVP, un module d'analyse du profil vocal, greffé sur MULTI-SPEECH ou CSL de Kay, ou encore de IVAS, greffé sur le logiciel d'analyse acoustique TFR de Avaaz. La seconde comprend les trop rares logiciels qui regroupent dans un même environnement ergonomique des programmes spécifiques aux principales applications cliniques des domaines se rapportant à la voix et la parole de la phoniatry, l'ORL et la neurologie, tels que DOCTOR SPEECH de Tiger Electronic ou PCLX de Laryngograph. Ce sont bien sûr les logiciels les mieux adaptés à nos préoccupations. Ce qui ne veut pas dire qu'ils sont entièrement satisfaisants, loin s'en faut. Ils n'ont pas tous les mêmes qualités. Certains sont plus aisés à utiliser que d'autres, par contre plus achevés au plan des mesures physiques ou d'un champ d'application plus complet. Très peu font l'effort de regrouper toutes les qualités demandées à de tels logiciels. Certains sont en train d'y arriver comme le système DIANA de SQ-Lab. Cela ne peut que favoriser et généraliser l'utilisation d'outils et de méthodes qui sont appelés à devenir incontournables dans l'avenir.

Largeur de bande et linéarité	Taux de distorsion harmonique	Rapport signal sur bruit	Pleurage	Séparation des canaux
30-13 kHz + - 3 dB	Magnétophone à cassette Hi-Fi standard SONY TC-RX361 2%	52 dB bande normale	0,40%	40 dB
25-17 kHz + - 3 dB	Magnétophone à cassette professionnel TASCAM 122 Mk III 1,2 %	58 dB bande normale	0,06%	42 dB
30-20 kHz + - 3 dB	Magnétophone à bande professionnel REVOX PR 99 Mk II 2% à 19 cm/sec	52 dB bande normale	0,08 %	45 dB
20-20 kHz + - 1 dB	Magnétophone numérique DAT portable SONY TCD-D10PRO II inférieur à 0,06 %	supérieur à 85 dB	0,01%	80 dB
20-20 kHz + - 0,5 dB	Magnétophone numérique DAT TASCAM DA-30 MkII inférieur à 0,00%	supérieur à 92 dB	inférieur à 0,00%	85 dB
20-20 kHz + - 3 dB	Magnétophone numérique baladeur MiniDisk SONY MZ-R35 inférieur à 0,01%	supérieur à 90 dB	Inférieur à 0,01%	85 dB
5-20 kHz + - 0,3 dB	Magnétophone numérique MiniDisk Desk SONY MDS-JB940 Inférieur à 0,00%	supérieur à 100 dB en lecture	Non mesurable	90 dB

**Tableau 2**  
*Caractéristiques comparées de quelques magnétophones à cassettes,  
à bande, numériques (DAT) et MiniDisk*

Largeur de bande et linéarité	Taux de distorsion harmonique	Rapport signal sur bruit	Commentaires
30-20 kHz + - 0,2 dB	Turtle Beach “ MONTEGO ” 0,00%	84 dB	Youpiiii !!!
30-17 kHz + - 0,2 dB	Ensonic AUDIO PCI 0,01%	74 dB	Ouais
40-19 kHz + - 0,12 dB	Creative Labs SOUNDBLASTER AWE64 GOLD 0,01%	71,5 dB	Mmmeu..
20-18 kHz + - 0,15 dB	Creative Labs SOUNBLASTER 16 CT 2940 0,04%	61 dB	Bof
40-11 kHz + - 3 dB	HiVal WAVE 96 0.04 %	44 dB	Beurk....

*On note de grandes différences. La meilleure n'est pas la plus chère.  
Par contre, la plus médiocre est bien la plus économique mais elle a les performances  
d'un mauvais magnétophone à cassettes exceptée la distorsion.*

**Tableau 3**  
*Caractéristiques comparées de quelques cartes audio pour PC*

### Références bibliographiques

- BAKEN, R. J. (1987). *Clinical measurement of speech and voice*. Taylor and Francis, London, p. 518.
- BAKER, K.K. ; RAMIG, L.O. ; LUSCHEI, E.S ; SMITH, M.E. (1998). Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging, *Neurology*, 51, p. 1592-1598.
- DEJONCKERE, P. & VILLAROSA, D. (1986). Analyse spectrale moyennée de la voix : Comparaison de voix normales et de voix altérées par différentes catégories de pathologies laryngées. *Acta oto-rhino-laryngologica bel.*, 40, p. 426-435.
- DEJONCKERE, P. (1990). Bruit de turbulence et apériodicité dans la voix pathologique : une approche multifactorielle. *Revue de Laryngologie*, 111, p. 353-357.

- DEJONCKERE, P. ; OBBENS, C. ; De MOOR, G. M. ; WIENEKE, G.H. (1993). Perceptual evaluation of dysphonia : Reliability and relevance. *Folia Phoniatrica*, 45, p. 76-83.
- DEJONCKERE, P. ; BRADYIER, P. ; CLEMENTE, P. ; CORNUT, G. ; CREVIER-BUCHMAN, L. ; FRIEDREICH, G. ; Van De HEYNING, P. ; REMACLE, M. ; WOIZARD, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur. Arch. Otorhinolaryngol.*, 258, p. 77-82.
- De KROM, G. (1994). Consistency and reliability of voice quality ratings for different type of speech fragments. *Journal of Speech and Hearing Research*, 37, p. 985-1000.
- De KROM, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowels fragments. *Journal of Speech and Hearing Research*, 38, p. 794-811.
- GENTIL, M. (1990). Acoustic characteristics of speech in Friedreich's disease, *Folia Phoniatrica et Logopaedica*, Karger, 42, p. 125-134.
- GIOVANNI, A. ; ROBERT, D. ; ESTUBLIER, N. & TESTON, B. (1996). Objective evaluation of dysphonia : Preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements, *Folia Phoniatrica et Logopaedica*, Karger, 48, p. 175-185.
- GIOVANNI, A. ; OUAKNINE, M. & TRIGLIA, J.M. (1999), Determination of largest Lyapounov exponents of vocal signal : Application to unilateral laryngeal paralysis, *Journal of Voice*, 13, p. 341-354.
- GIOVANNI, A., HEIM, C. ; DEMOLIN, D. & TRIGLIA, J. M. (2000). Estimated subglottic pressure in normal and dysphonic subjects, *Ann. Otol. Rhinol. Laryngol*, 109, p. 500-504.
- HAMMARBERG, B. ; FRITZELL, B. ; GAUFFIN, J. ; SUNDBERG, J. & WEDIN, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, vol. 90, p. 441-451.
- HIRANO, M. ; KOIKE, Y. & Von LEDEN, H. (1968). Maximum phonation time and air usage during phonation. *Folia Phoniatrica*, 20, p. 185-201.
- HIRANO, M. (1981). Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice. *Clinical Evaluation of Voice*, Springer Verlag, Wien.
- HIRANO, M. (1989). Objective evaluation of the human voice: Clinical aspects, *Folia Phoniatrica*, vol. 41, p. 89-144.
- HIRAOKA, N. ; KITAZOE, Y. & UETA, H. (1984). Harmonic-intensity analysis of normal and hoarse voices. *Journal of the Acoustical Society of America*, 76, p. 1648-1651.
- JIANG, J. ; O'MARA, T. ; CHEN, H. J. ; STERN, J.I. ; VIAGOS, D. & HANSON, D. (1999). Aerodynamic measurements of patients with Parkinson's disease. *Journal of Voice*, 13, p. 583-591.
- KASUYA, H. ; OGAWA, S. ; MASHIMA, K. & EBIHARA, S. (1986). Normalised noise energy as an acoustic measure to evaluate pathologic voice. *Journal of the Acoustical Society of America*, 80, p. 1329-1334.
- KREIMAN, J. ; GERRATT, B.R. ; KEMPSTER, G.B. ; ERMAN, A. & BERKE, G.S. (1996). Perceptual Evaluation of Voice Quality : Review, Tutorial, and a Framework for Future Research, *Journal of Speech and Hearing Research*, vol. 36, p. 21-40.



- LUCKHAUPT, H. (1997). The history of laryngology in Europe. In O., Kleinsasser, H., Glanz and J., Olofsson (eds), *Advances in Laryngology in Europe*, Elsevier, Amsterdam, xvii-xxxiii.
- PEREZ, K.S. ; RAMIG, L.O. ; SMITH, M., E. & DROMEY, C. (1996). The Parkinson larynx : Tremor and videostroboscopic findings. *Journal of Voice*, 4, p. 354-351.
- PLANT, R. & HILLEL, A. (1998). Direct measurement of subglottic pressure and laryngeal resistance in normal subject and spasmodic dysphonia. *Journal of Voice*, 12, p. 300-314.
- PICCIRELLO, J. ; COLIN, P. ; DENNIS, F. & FREDERICKSON, J. (1998). Multivariate analysis of objective vocal function. *Ann. Otol. Rhinol. Laryngol.*, 107, p. 107-112.
- QI, Y. ; WEINBERG, B. ; BI, N. & HESS, W.K. (1995). Minimizing the effect of period determination of the computation of amplitude perturbation in voice. In D., Wong (ed.), *Workshop on acoustic voice analysis*, Iowa City, IA: National Center for voice and speech.
- RAMIG, L. & SHIPP, T. (1987). Comparative measures of vocal tremor and vocal vibrato. *Journal of Voice*, 12, p. 162-167.
- REVIS, J. ; GIOVANNI, A. ; WUYTS F.L. & TRIGLIA, J.M., (1999). Comparison of different voice samples for perceptual analysis, *Folia Phoniatr. Logop.*, 51, p. 108-116.
- SCHUTTE, H. (1980). *The efficiency of voice production*. Kemper, Groningen.
- SMITHERAN, J. & HIXON, T., A. (1981). A clinical method for estimating laryngeal airway resistance during vowels production. *Journal of Speech and Hearing Disorders*, 46, p. 138-148.
- STELZIG, Y. ; HOCHHAUS, W. ; GALL, V. & HENNEBERG, A. (1999). Laryngeal manifestations in patients with Parkinson disease, *Laryngo-Rhino-Otologie*, 8, p. 544-551.
- SULTER, A.M. ; WIT, H.P. ; SCHUTTE, H.K. & MILLER, D., G., (1994). A structure approach to voice range profile (Phonetogram) analysis. *Journal of Speech and Hearing Research*, 37, p. 1076-1085.
- TESTON, B. ; GALINDO, B., (1995). A diagnostic and rehabilitation aid workstation for speech and voice pathologies, *Proceedings of Eurospeech 4*, European. Speech Communication Association, Madrid, p. 1883-1886.
- TITZE, I.R. (1992). Acoustic interpretation of the voice range profile. *Journal of Speech and Hearing Research*, 35, p. 21-34.
- TITZE, I.R. (1995). Summary statement. *Workshop on acoustic voice analysis*. Iowa City: National center of voice and speech, 36.
- WUYTS, F.L. ; De BODT, M.S. ; MOLENBERGHS, G. ; REMACLE, M. ; HEYLEN, L. ; MILLET, B. ; LIERDE, K.V. ; RAES, J. & Van De HEYNING, P.H. (2000). The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech and Hearing Research*, 43, p. 796-809.
- YU, P. (2001). *Méthodes instrumentales d'analyse de la dysphonie : corrélation avec l'analyse perceptive*, Thèse de l'université de la Méditerranée, Aix-Marseille II, Laboratoire d'Audio-Phonologie clinique, 159.
- YUMOTO, E. & GOULD, W.J. (1982). Harmonics to noise ratio as an index of the degree of hoarseness, *Journal of the Acoustical Society of America*, vol. 71-6, p. 1544-1550.